

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

---

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**





No.	Score	Match	Length	DB	ID	Description
1	578.2	25.9	616	61	AL045824	AL045824 DKFZP434H
2	477.2	21.3	512	79	AM300030	AM300030 xs55h01.x
3	457.8	20.5	473	44	AI269960	AI269960 qk18d11.x
4	440.4	19.7	485	39	AA660926	AA660926 ak22d06.s
5	437.6	19.6	467	51	AI761354	AI761354 wif6e08.x
6	435.4	19.5	464	51	AI739238	AI739238 w14907.x
7	423	18.9	473	63	AI984995	AI984995 wr9906.x
8	405.6	18.1	436	26	W22160	W22160 63A6 Human
9	403	18.0	411	49	AI628985	AI628985 cy79c02.x
10	394.2	18.0	411	49	AI629007	AI629007 ty79e02.x
11	394.2	18.0	411	49	AI629007	AI629007 ty79e02.x
12	372.8	16.7	384	60	AI796741	AI796741 wa14d02.x
13	371.2	16.6	385	44	AI722081	AI722081 qj78f01.x
14	371.2	16.6	385	44	AI722081	AI722081 qj78f01.x
15	369.2	16.5	386	46	AI453367	AI453367 t3j7c11.x
16	355.2	15.9	360	63	AI991419	AI991419 ws10b04.x
17	351.4	15.7	367	79	AM304132	AM304132 xs13f01.x
18	341.4	15.3	381	80	AM351654	AM351654 QV2-CT014
19	335.6	15.0	342	51	AI738967	AI738967 w112f02.x
20	334	14.9	342	51	AI738967	AI738967 w112f02.x
21	304.6	13.6	311	35	AA551799	AA551799 nk04a11.s
22	292.2	13.1	317	32	AA344913	AA344913 EST50856
23	290.4	13.0	335	28	AA121174	AA121174 z188g08.s
24	287.8	12.9	316	74	AM192301	AM192301 x184a12.x
25	270.4	12.1	272	44	AI270726	AI270726 q193d07.x
26	267.4	12.0	289	63	AA348243	AA348243 EST54707
27	266.4	11.9	278	63	AM013904	AM013904 UI-H-B10
28	248.8	11.1	279	61	AI869919	AI869919 we75h03.x
29	240.8	10.8	244	60	AI783811	AI783811 lu79c11.x
30	197.4	8.8	216	61	AI869907	AI869907 we75g03.x
31	180.8	8.1	206	31	AA327309	AA327309 EST30621
32	160	7.2	204	28	AA121198	AA121198 z188g08.r
33	143	6.4	353	49	AI642234	AI642234 vn68c05.y
34	131.2	5.9	221	66	AM062449	AM062449 MR0-CT006
35	120.6	5.4	473	36	AA616807	AA616807 vn68c05.r
36	105	4.7	463	105	AQ320634	AQ320634 RPTC11-92
37	68.2	3.1	823	84	B03396	B03396 cSRL-17BD10
38	67.2	3.0	443	102	AQ100902	AQ100902 HS-3060_B
39	61.6	2.8	997	82	CNS005TE	AI060767 Drosophila
40	57.2	2.6	587	36	C79507	C79507 C79507 Mous
41	56.6	2.5	571	92	AQ919110	AQ919110 RPTC1-2
42	55.6	2.5	476	36	C79500	C79500 C79500 Mous
43	55.6	2.5	445	34	AA464787	AA464787 zxb8g10.r
44	55.2	2.5	498	50	AI676456	AI676456 etmsr021
45	54.6	2.4	657	80	AM369742	AM369742 IL0-BT015

ALIGNMENTS

RESULT 1  
LOCUS AL045824/c  
DEFINITION DKFZP434H36\_s1 434 (synonym: htes3) Homo sapiens CDNA clone  
ACCESSION AL045824  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 616)  
AUTHORS Koehler, K., Beyer, A., Wewes, H. W., Gassenhuber, J. and Wiemann, S.  
TITLE EST (Koehler, et al.)  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3137188.  
CONTACT: Koehler K  
MISF  
Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 3' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
ri sequence also available.  
This clone is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
Location/Qualifiers  
1. 616  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZP434H36"  
/clone\_lib="434 (synonym: htes3)"  
/tissue\_type="testis"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Vector: pSPort1; Site\_1: NotI; Site\_2: SalI"

BASE COUNT 134 a 142 c 185 g 155 t

Query Match 25.9%; Score 578.2; DB 61; Length 616;  
Best Local Similarity 98.4%; Pred. No. 1; 3e-134;  
Matches 605; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY	1603	atggcgaataaagggaaggttgtagagactaaccttcgtagagctgacgtgacgtg	1662
DB	613	atggcgaataaagggaaggttgtagagactaaccttcgtagagctgacgtgacgtg	554
QY	1663	cagaagccttggaataaaggcgagcgtgacgtgacgtgacgtgacgtgacgtgac	1722
DB	553	cagaagccttggaataaaggcgagcgtgacgtgacgtgacgtgacgtgacgtgac	494
QY	1723	aaggagatgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac	1782
DB	493	aaggagatgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac	435
QY	1783	gggttagaagaacgtgagcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtg	1842
DB	434	-ggttagaagaacgtgagcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtg	376
QY	1843	gagagagcacaatgac	1902
DB	375	gagagagcacaatgac	316
QY	1903	aatctcaactccttggtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtg	1962
DB	315	aatctcaactccttggtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtg	256
QY	1963	tgtgaagcacaatgag	2022
DB	255	tgtgaagcacaatgag	196
QY	2023	ccagatccatccttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	2082
DB	195	ccagatccatccttcttcttcttcttcttcttcttcttcttcttcttcttcttctt	136
QY	2083	tcttcttcttgaacccaataaagaacccaataaagaacccaataaagaacccaataa	2142
DB	135	tcttcttcttgaacccaataaagaacccaataaagaacccaataaagaacccaataa	76
QY	2143	ggcctctgctcaatattgctgagagattgacactcaaaccttacttgagctgacatt	2202
DB	75	ggcctctgctcaatattgctgagagattgacactcaaaccttacttgagctgacatt	16
QY	2203	aaaataaagacatt 2217	
DB	15	aaaataaagacatt 1	

RESULT 2  
AM300030/c



LOCUS AM300030 512 bp mRNA EST 18-JAN-2000  
DEFINITION xs55h01.x1 NCI\_CGAP\_kid11 Homo sapiens cDNA clone IMAGE:2773585 3',  
ACCESSION AM300030  
VERSION AM300030.1 GI:6709707  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 512)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
TITLE Unpublished (1997)  
JOURNAL On May 18, 1998 this sequence version replaced gi:3137941.  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bdrrp/image/image.html](http://www.bio.llnl.gov/bdrrp/image/image.html)

Seq primer: -400P from Gibco  
High quality sequence stop: 462.  
Location/Qualifiers  
1..512  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2773585"  
/clone\_1lb="NCI\_CGAP\_kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pRT3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 117 a 105 c 158 g 132 t  
ORIGIN

Query Match 21.3%; Score 477.2; DB 79; Length 512;  
Best Local Similarity 98.1%; Pred. No. 2.7e-109;  
Matches 504; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY 1712 tctgcccccaagagatgatgaagatgacctcttctgctgaagtccaaggaggaa 1771  
|||||  
DB 512 TCTGCCCCCAAGAGATGATGACATGAGCTGACCTCTT-CTGAGTCCAAAAGGGGAA 454  
|||||  
QY 1772 accaatcagcgcttggaagacagtagctcgcgcccaacctctgtaacacaagcctc 1831  
|||||  
DB 453 ACCAAATTCAC-CCTTGTGGAACAGTAGAGCTCCGGCCCACTGTGTAAACACAAAGCCTC 395  
|||||  
QY 1832 ggaacacgcttgagagagccacatgacacacacagatgcatcctctggacctgaatc 1891  
|||||  
DB 394 GGATGACCTTGAAGAGAGCGACACTACACACACAGATGGCAGTCTTGGGACCTGAATC 335  
|||||  
QY 1892 tatcaccacgaagatcctaactcccttggccttgaaccagggccagataaggaacgct 1951  
|||||  
DB 334 TATCACCACGAGAACTCAAACTCCCTTGGCCCTGAACACAGGCGCAGATGAAGAACAGCT 275  
|||||  
QY 1952 cgggcccattttggaaggccaatgtgagagaaaggagagcagccgctttggagaag 2011  
|||||

DB 274 CGGGCACACTTCTTGAAAGGCCACAGCTGGAGGAAGGAGCGCCAGCCACTTTGGAGAG 215  
QY 2012 atcccaaggatccagactcattcctctctctcctcctgcccagtgattgtctcccaag 2071  
|||||  
DB 214 ATCTCAAGGATCCAGACTCTCATCTTCTCTCGGCCACAGTGAATTTGGCTCTCCAG 155  
QY 2072 cttggggagactcctctcttgaaccctaataagaccacagtgagctctctctccat 2131  
|||||  
DB 154 CTCTGGGGAGACTCTTCTCTGAAACCCCTTAATGAAGCCCACTGGAGTCTCTCTCCAT 95  
QY 2132 cccctcctctgcccctctgtcttaattgctgcccagatgtcactccaacctactctg 2191  
|||||  
DB 94 CCTCTCTCCTCGCCCTCTGCTTAATTCGTCGCCAGATTTGTCACCTCAACCTTACTG 35  
QY 2192 agctcattataaataacagattatttcca 2225  
|||||  
DB 34 AGCTCATTAATTAATAATGAAGATTTATTTTCCA 1

RESULT 3  
AI269960/c 473 bp mRNA EST 28-JAN-1999  
LOCUS qk18d11.x1 NCI\_CGAP\_kid3 Homo sapiens cDNA clone IMAGE:1869333 3'  
DEFINITION similar to TR:Q14160 Q14160 KIA0147 PROTEIN ;, mRNA sequence.  
ACCESSION AI269960  
VERSION AI269960.1 GI:3889127  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 473)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bdrrp/image/image.html](http://www.bio.llnl.gov/bdrrp/image/image.html)

Insert Length: 520 Std Error: 0.00  
Seq primer: -400P from Gibco  
High quality sequence stop: 451.  
Location/Qualifiers  
1..473  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1869333"  
/clone\_1lb="NCI\_CGAP\_kid3"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pRT3D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pRT3D vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 92 a 159 c 118 g 104 t  
ORIGIN

Query Match 20.5%; Score 457.8; DB 44; Length 473;  
Best Local Similarity 99.6%; Pred. No. 1.9e-104;

Matches 459; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 654 tcagttgtgtcgaatcctgaggcgctgcagagcagctggctccctcgaaatcggga 713
DB 461 TCAGTTGTGTGGAATCTGGGGGCTGCGAGGACAGCTGGGCTCCCTTGGAAATCGGGA 402
OY 714 aaacagaagaagaagcttcacacagctgtgtagctcccgagccttgctgcagcat 773
DB 401 AAACAAGAGAAGAAGCTCTTATCATGAGCTGAGGCTCCCGAGGCTTGGCTGACAGCAT 342
OY 774 ttccagggccccatccagaaagcctgtgcatcttcatcagcattgtgaaacctgtccct 833
DB 341 TTCACAGGCGCCCATCCAAAGAGCTGGCATCTTTATCAGCCATGTAACCTGGCTCCCT 282
OY 834 gctctgtgaggtgaggttgagataaggagacagattgtcgaagtcaatgagcgtgacct 893
DB 281 GCTCTGTGAGGTGGAGTTGAGATAGGGAGCCGATTTGCAAGTCAATGGCTGCACCTT 222
OY 894 ccttaacctgtatcacaagaaggcgtgtaaatgtctgaaataatagcgcagacctgacat 953
DB 221 CTTACTACCTGGATACACAGAAGAGCTGTAATGTGCTGAAGAGTAGCCGACGCTGACAT 162
OY 954 ctccattgtagctgcagctgtgcgcggagctgttcatatgacagacggagcgctgtgcaga 1013
DB 161 CTCCTATTGTAGCTGCGAGCTGGCGGAGCGTGTTCATGACAGACCGGAGCGGCTGGCAGA 102
OY 1014 ggcgcgagcagcgtgagctcagcgagcagagcttcctcatcgcaagaacggctgagatga 1073
DB 101 GCGCGCGGAGCTGAGCTGACGCGGAGGAGAGCTTCTCATGCAAGACGGCTGGCGATGGA 42
OY 1074 gtccacaagaatcctcccaaggagcagcagagatgagcagc 1114
DB 41 GTCCACAAGATCTCTCCAGAGAGCAGAGATGAGAGCGGC 1
```

RESULT 4  
LOCUS AA860926/c 485 bp mRNA EST 04-JAN-1999  
DEFINITION ak22406.s1 Soares testis\_NHT Homo sapiens cDNA clone IMAGE:1406699  
ACCESSION AA860926  
VERSION AA860926.1 GI:2953066  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 485)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2287585.  
CONTACT: Robert Strausberg, Ph.D.  
Email: (301) 496-1550  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLNL at:  
www.bio.lnl.gov/bbrp/image/image.html

## FEATURES

source 1. 485  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1406699"  
/clone\_lib="Soares\_testis\_NHT"  
Location/Qualifiers  
Insert Length: 629 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence scop: 426.

/sex="male"  
/lab\_host="DH10B"  
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified  
polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
was prepared from mRNA obtained from Clontech  
Laboratories, Inc., and primed with a Not I - oligo(dT)  
primer [5',  
TGTTCACATCTGAGAGTGAGAGCGGCCGCCCAATTTTTTTTTTTT 3'].  
Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library  
went through one round of normalization to Cot5, and was  
constructed by Bento Soares and M. Fatima Bonaldo. "  
BASE COUNT 112 a 98 c 149 g 126 t  
ORIGIN

Query Match 19.7%; Score 440.4; DB 39; Length 485;  
Best Local Similarity 98.1%; Pred. No. 4.5e-100;  
Matches 477; Conservative 0; Mismatches 6; Indels 3; Gaps 3;

```
OY 1740 gctgacctctgtcgtgaagtcacaaagggaacacaaatcacggttgagaaacagtgta 1799
DB 485 GCTGACCTTCTT-CTGAAGTCCAAAAGGGGAAACCAAAATTCAC-CGTTAGGAAAGATGA 428
OY 1800 gctcggccccacactcgtgtaacacaaagcctcgagcagccttgagagagcgcaatgac 1859
DB 427 GCTCGGCCCCACACTCTGTATACCAAAAGCCTGAGACAGCCTTGAGAGAGGCGCACACTAC 368
OY 1860 aacacacagaatgacatccttggagctgaatctatcaccaggaatctcaacatccctt 1919
DB 367 ACACACCAATGAGCATCTCTTGGAGACTGATATCCACAGAAATCTCAAACTCCCTTT 308
OY 1920 ggcgcctgaacacagggaacaaatagaaagcgtcggagcacttttgaaggccaatgtag 1979
DB 307 AG-CTGAACCAAGGCGCAATAGGAACAGCTCGGGCCACTCTTCTGAAGGCCAATGTGG 249
OY 1980 aggaagaaggagcagcagcagcgtttggaagaagatctcaagatccagactccatccctt 2039
DB 248 AGGAAGAAGGAGAGCAGCAGCGCTTTGGAGAAAGATCTCAAGATCTCAGATCTCATTTCTT 189
OY 2040 tccctctgcccagtgaaattgtctctcccaagcttggaggacatccctcttgaaccccta 2099
DB 188 TCCTCTGGCCCAATGAAATTGTGCTCTCCACACTCTGGGGAGACTCCTTCTTGAACCTTA 129
OY 2100 ataagaaccccatgagatctctctctccatccctctctctgccccttgcttcaatgt 2159
DB 128 ATTAAGACCCCACTGAGATCTCTCTCTCCATCCCTCTCTGCGCTGCTTAATTG 69
OY 2160 ctgcagagattgtcactccaacacttactctgagctcatataataaacaagattat 2219
DB 68 CTGCCAGAGATTGTCTACTCCAAACCTTACTCTGAGCTCATTAATAATAACAGATTAT 9
OY 2220 ttcca 2225
DB 8 TTTCGA 3
```

RESULT 5  
LOCUS A1761354/c 467 bp mRNA EST 24-JUN-1999  
DEFINITION w160e08.x1 NCI-CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2394662 3',  
RNA sequence.  
ACCESSION A1761354  
VERSION A1761354.1 GI:5177021  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 467)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

JOURNAL  
COMMENT

Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3138298.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.1nl.gov/bdrp/image/image.html

FEATURES  
Source

Seq primer: -400P from Glibco.  
Location/Qualifiers  
1. 467  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 108 a 96 c 143 g 120 t  
ORIGIN

Query Match 19.5%; Score 437.6; DB 51; Length 467;  
Best Local Similarity 98.7%; Pred. No. 2.2e-99;  
Matches 462; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1765 aggggaacacaaatcagcgcgttagaagacagtcgcgcgcacacccctgtgaacaca 1824  
DB 467 AGGGGAACCAAAATTCAC-CGTTAGGAACAGTAGCTCCGCCCCACCTCGTAACACA 409  
QY 1825 aagcctcgagcagccttgagagggcacatgacacacacagatgcatccttggac 1884  
DB 408 AAGCCTCGAGCACGCTTGAAGAGGCGCACACTACACACACAGATGGCATCTTGGGAC 349  
QY 1885 ctgaatctatcacccagaatctcaactcccttggccttgaaacagggccagataag 1944  
DB 348 CTGAATCTATCACCAAGAAATCTCAAACTCCCTTGGCCCTGGAACCAAGGCAATTAAG 289  
QY 1945 aacagctcgagccact-ttttgaagccaaatgtgaggaaggaagcagcagcgtt 2003  
DB 288 AACAGCTCGGGCCACTCTTTTGAAGCCAAATGGAAGAAAGGACACAGCCGTTT 229  
QY 2004 ggggaagatctcaagatccagatctccttccctcttgccagtgaaattggc 2063  
DB 228 GGGAGAGATCTCAAGGATCCAGACTCTCATTCCTTCTCTGCGCCACGTAATTTGGTC 169  
QY 2064 tctccagccttgggggaaccccttccttgaaacctataagaccacagatcctct 2123  
DB 168 TCTCCAGCTCTGGGAGACTCTTCTTGAACCCCTAATAAGACCCCACTGAGCTCTCT 109  
QY 2124 ctctccatccctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2183  
DB 108 CTCTCCATCCCTCTCCCTCTGCGCTCTCAATTTGCTGCCAGAGATGTCTACTCCAAAC 49  
QY 2184 tctactgagcctaatataaataaagattatttccagcttaa 2231  
DB 48 TTACTGTAGCTCATTAATAAATAACAGATTATTTTCAGCTAAA 1

RESULT 6  
LOCUS A1739238/c  
DEFINITION w114407.x1 NCI-CGAP\_Col6 Homo sapiens cDNA clone IMAGE:2390268 3',  
mRNA sequence.  
ACCESSION A1739238  
VERSION A1739238.1 GI:5101219  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 464)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:31389064.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
www-bio.1nl.gov/bdrp/image/image.html

FEATURES  
Source  
1. 464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 104 a 90 c 140 g 130 t  
ORIGIN

Query Match 19.5%; Score 435.4; DB 51; Length 464;  
Best Local Similarity 98.7%; Pred. No. 7.8e-99;  
Matches 439; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1792 aacgtgagctcggcccccacccctctgtaacaaagcctcgaccagccttgagagagc 1851  
DB 464 AACGTAGCTCCGCCCCACCTCTGTAAACAAAGCCTCGACCACTTGAAGAGGC 405  
QY 1852 cacatgacacacacacagatgcatccttggagcctgatatcaccagagatcctaa 1911  
DB 404 CACACTACACACACACAGATGGCATCTTGGAGCTGATGATCAACCCAGGAATCTCAA 345  
QY 1912 ctcccttggccttgagacccagggccagataaggaagcctcgagcacttttgaagc 1971  
DB 344 CTCCCTTTGGCCCTGGAACCAAGGCGCAGATTAAGAAAGCTCGGCCACTCTTCTGAAGGC 285  
QY 1972 caatgtgaggaagagagcagcagccttgggaagatctcaagatccagactct 2031

Db 284 CAATGTGAGGAAGGAGCAGCAGCGTTTGGAGAAATCTCAAGATCCAGACTCT 225  
Qy 2032 ccttccttcctcctgagccagatgaattgtctcctccagcttggaggaccttcctt 2091  
Db 224 CATTCCTTCTCTGAGCCAGTAATTTGGTCTCTCCAGCTCTGGGGAGACTCTCTT 165  
Qy 2092 gaaccctaaagaagcccaactgagctctctctcctccatccctcctctg 2151  
Db 164 GAACCTTAATAGACCCAGCTGAGTCTCTCTCCATCCCTCTGCTGCGCTCTG 105  
Qy 2152 tctaatgtcgcagagatgtcaactcaactctctgagctctctataataaac 2211  
Db 104 TCTAATGTGTCAGAGATTGTCACTCCAAACCTTACTGTGAGCTCATTAATAAATTAAC 45  
Qy 2212 agattatttccagcttaaaaaa 2236  
Db 44 ACATTTATTTTCCAGCTAAAAAAA 20

RESULT 7  
A1984995/c 473 bp mRNA EST 31-AUG-1999  
LOCUS wr89g06.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2494906 3',  
DEFINITION mRNA sequence.  
ACCESSION A1984995  
VERSION A1984995.1 GI:5812272  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 473)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 18, 1998 this sequence version replaced gi:3138171.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. consortium/ILN at:  
www.bio.lnlnl.gov/bdrp/image/image.html

FEATURES  
source  
Seq primer: -40UP from Gibco  
High quality sequence stop: 463.  
Location/Qualifiers  
1..473  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2494906"  
/clone\_lib="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not 1; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneIDs 132376-132391, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 110 a 96 c 145 g 121 t 1 others  
ORIGIN  
Query Match 18.9%; Score 423; DB 63; Length 473;

Best Local Similarity 96.0%; Pred. No. 1e-95;  
Matches 455; Conservative 0; Mismatches 16; Indels 3; Gaps 2;  
Qy 1751 tctgtagtccaaaagaggaaacccaattcaacgcttgaagaacgttgctccgcccc 1810  
Db 471 TTTCTGAATGTCAAAAGGGGAACCAAAATTCAC-CGTTAGGAAACAGTGAAGTCCGGCCCC 413  
Qy 1811 acctctgaaacaaagctcggaccagcttgaagagagacatgaaacaaacagat 1870  
Db 412 ACCCTGTGAACACANAGCTCCGATC--AGCTTAGAGAGAGCCACATACACACAGAT 355  
Qy 1871 ggaatccttggagacctgaatcctatacaccaggaatcctaaactccttggccctgaac 1930  
Db 354 GGCATCTCTTGGGACCTGGAATTCATCACCCAGGAATCTCAAACTCCCTTGGCCCTGAAC 295  
Qy 1931 agggccagataaggaacacatcggccacttttgaagaagcaalytgaaggaagagag 1990  
Db 294 AGGGCCAGATAAGGAACACCTGGGCCACTCTTCTGAAGGCCAACCTGGAGGAAGGGAG 235  
Qy 1991 cagccagcgtttggagagagatctcaagatccagactcatctcctctgccc 2050  
Db 234 CAGCCAGCCATTTGGAGAGAGATCTCAAGATCCAGACTCATCTTCTCTGCGCCC 175  
Qy 2051 agtgaattgtctctcctcagctcttggagactcctccttgaacccataagacc 2110  
Db 174 AGTAATTTGGTCTCTCCAGCTCTGGGAGACTCTTCTTGAACCCCTAATTAAGACCCGA 115  
Qy 2111 ctggagctcctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2170  
Db 114 CTGGAATCTCTCTCTCCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 55  
Qy 2171 gtcaaccacaaactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 2224  
Db 54 GTCACCTCAAACTTACTCTGAGCTCATTAATAATAACAGATTATTTTCC 1

RESULT 8  
W22160 636 bp mRNA EST 06-MAY-1996  
LOCUS W22160  
DEFINITION 63A6 Human retina cDNA Tsp509T-cleaved sublibrary Homo sapiens cDNA  
not directional, mRNA sequence.  
ACCESSION W22160  
VERSION W22160.1 GI:1299007  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 636)  
AUTHORS Macke, J., Smallwood, P. and Nathans, J.  
TITLE Adult Human Retina cDNA  
JOURNAL Unpublished (1996)  
COMMENT On Apr 14, 1993 this sequence version replaced gi:785820.  
Contact: Dr. Jeremy Nathans  
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics  
Johns Hopkins School of Medicine  
725 North Wolfe Street, Baltimore, MD 21205  
Tel: 410 955 4678  
Fax: 410 614 0827  
Email: jeremy.nathans@jhu.edu  
Clones from this library are NOT available.  
PCR Primers  
FORWARD: CTTTGGACAGTTCAGCTCGGTTAAGT  
BACKWARD: GAGGTGCTTATGATGATTTCTTCCAGGGTAA  
Seq primer: GGGTAAAGCAAAAGATT.  
Location/Qualifiers  
1..636  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Human retina cDNA Tsp509T-cleaved sublibrary"  
/sex="mixed (males and females)"  
/tissue\_type="retina"  
/dev\_stage="adult"



REFERENCE 1 (bases 1 to 411)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 16, 1998 this sequence version replaced gi:2961764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

FEATURES  
source  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..411  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2285306"  
/clone\_1lb="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: Kidney; Vector: p773D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(clonoids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 95 a 83 c 131 g 102 t  
ORIGIN

Query Match 18.0%; Score 403; DB 49; Length 411;  
Best Local Similarity 98.8%; Pred. No. 1e-90;  
Matches 406; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1808 cccactctgtgaacacaaagcctcgagccttgagagagggccacatgacacacacca 1867  
|||||  
411 CCCACCTCGTAACACAAAGCCTCGGACCGCTTGAGAGGCCACACTACACACACCA 352  
1868 gatgcatccttggagacccaatctatccacaggaatcacaactcccttggccctga 1927  
|||||  
351 GATGGATCTTGGGACCTGGAATCTATCACCCAGAAATCTCAAACTCCCTTGGCCCTGA 292  
1928 accagggcagataaggaacagctcgagcacttttgaagggccaatgtgaggaag 1987  
|||||  
291 ACCAGGGCCAGATAAGGAACAGCTCGGGCCACCTCTTAAGGCCAATGTGAGGAAGG 232  
1988 gaggcagccagccgcttggaggaagatcctaagatccaactctatcttccctctgg 2047  
|||||  
231 GAGCACCCAGCGTTGGGAGAGATCTCAAGATCCAAGATCTCAATCTCTTCCTCTGG 172  
2048 ccagatgaatttgctctccagccttggaggaactccttcccttgaacctataagacc 2107  
|||||  
171 CCCAGGAATTTGGTCTCTCCACAGCTCTGGGGAGCTCTTCTCTGAACCTTAATAAGACC 112  
2108 ccactgagatctctctctcctcctcctctcctctcctctgccccttgctctaaattgctgcag 2167  
|||||  
111 CCATGGGAATCTCTCTCTCCATCCCTCTGCTGCTGCTGCTTAATGCTGTCAGG 52  
2168 attgtcaactcaaacctactctgtagactcaataataaacagactta 2218  
|||||  
51 ATTGTCACTCCAAACCTTACTCTGAGCTCATTAATAATAAACAAGATTTA 1

RESULT 11  
LOCUS AI989344/c  
DEFINITION w227e03.x1 NCI\_CGAP\_Brn53 Homo sapiens cDNA clone IMAGE:2559292 3',  
mRNA sequence.  
ACCESSION AI989344  
VERSION AI989344.1 GI:5836225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 407)  
AUTHORS NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BRGAP), Tumor Gene Index  
JOURNAL Unpublished (1998)  
COMMENT On Feb 18, 1999 this sequence version replaced gi:4296517.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

FEATURES  
source  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..407  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2559292"  
/clone\_1lb="NCI\_CGAP\_Brn53"  
/tissue\_type="three pooled meningiomas"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: PCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies."

BASE COUNT 97 a 81 c 127 g 102 t  
ORIGIN

Query Match 17.6%; Score 394.2; DB 63; Length 407;  
Best Local Similarity 98.0%; Pred. No. 1.6e-88;  
Matches 399; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

1818 gaacacaaagcctcgagcacttgagagagggccacatgacacacacagatgcatcc 1877  
|||||  
407 GAACACAAAGCCTCGGATCAGCTTGAGAGAGGCCACACTACACACAGATGGCATCC 348  
1878 ttggagcctggaatctataccccaaggaatcctaactcccttggcccttaaacaggagca 1937  
|||||  
347 TTGGGACCTGAATCTATCACCCAGGAATCTCAAACTCCCTTGGCCCTGAACCAAGGCCA 288  
1938 gataaagaacagctcgagcacttttgaagggccaatgtgaggaaggaaggaacagcag 1997  
|||||  
287 GATAAGAACAGCTCGGGCCACTCTCTGAAGGCCAAGCTGAGAGGAAGGAGACAGCCAG 228  
1998 ccgttggaggaagatcctaaggaatcctaactcttccctctgcccagtgaaat 2057  
|||||  
227 CCATTGGGAGAGATCTCAAGATCCAAGATCCAAGATCTCAATCTCTTCCCTGCGCCAGTAAT 168  
2058 ttggtctctccagccttggaggaactccttcccttgaacctataagacccacaggaat 2117  
|||||  
167 TTGGTCTCTCCAGGCTCTGGGGAGCTCTTCTTGAACCTTAATAAGACCCACATGGAAT 108  
2118 cctctctcctcctcctctcctctgcccctctgctcttaattgtgcaggaattgtcactc 2177

	RESULT	12	
	A1796741/c		
	LOCUS	A1796741	384 bp mRNA EST 16-DEC-1999
	DEFINITION	wal4d02.x1 NCI_CGAP_Kid11 Homo sapiens CDNA clone IMAGE:2298051 3'	
	ACCESSION	A1796741	mRNA sequence.
	VERSION	A1796741.1	GI:5362204
	KEYWORDS	EST.	
	SOURCE	human.	
	ORGANISM	Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
		Eutheria; Primates; Catarrhini; Homnidae; Homo.	
	REFERENCE	1 (bases 1 to 384)	
	AUTHORS	NCI-CGAP htcp://www.ncbi.nlm.nih.gov/ncicgap.	
	TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
	JOURNAL	Unpublished (1997)	
	COMMENT	On May 18, 1998 this sequence version replaced gi:3138695. Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LDNW at: <a href="http://www-bio.linn.gov/dbrrp/image/image.html">www-bio.linn.gov/dbrrp/image/image.html</a>	
	FEATURES	Insert Length: 496 Std Error: 0.00 Seq primer: -40UP from Gibco.	
	source	Location/Qualifiers	
		1..384	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:2298051"	
		/clone_lib="NCI_CGAP_Kid11"	
		/lab_host="DH10B"	
		/note="Organ: Kidney; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; Plasmid DNA from the normalized library NCI_CGAP Kid3 was prepared, and ss circles were made in vitro. Following HAP hybridization, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneds 1323376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."	
	BASE COUNT	95 a 75 c 120 g 94 t	
	ORIGIN		
	Query Match	16.7%; Score 372.8; DB 60; Length 384;	
	Best Local Similarity	98.2%; Pred. No.3.7e-83;	
	Matches 377; Conservative	0; Mismatches 7; Indels 0; Gaps 0;	
Oy	1846	agaagccacatgacacacacacgatgcattccttggagactgaatcatcacccgaagt 1905	
Db	384	AGAGGCCACACTACACACACCAGATGGCATCTTGAGCACTGAATCTATCACCCAGAAT 325	
Oy	1906	ctcaactccctttggacctaaacaggcgccagatgaagaagaagcttgccactttt 1965	
Db	324	CTCAAAGTCCTTTTGCCCCGTGAACCAAGGGCCAGATTAAGAAACAGCTCGGGCCACTTCT 265	

Oy	1966	gaagggccaatcttgaggagaaggaaacgaacgacgcgtttggagaagaatcctaagatcca	2025
Db	264	GAAGGCCAACGCTGGAGAGAAAGGACACACCACCATTTTGGAGAAGAATCTCAAGATGCCA	205
Oy	2026	gactctcatcttccttctctcctcgcccagtagtaatttgtctctctccagcctttggyggaactcc	2085
Db	204	GACTCTCATTCCTTCCCTCGTG6CCACGAGTAATTGGTCTCTCCAGCCTCTGGGGACATCC	145
Oy	2086	tctcttgaacctataaagaccgccacttggagtctctctctctccatccctctctctctgc	2145
Db	144	TTCCTTGAACCTTAATAAAGACCCCATGTGAGTCTCTCTCTCCATCCCTCTCTGC	85
Oy	2146	ctctgctctaattgtctgcgaagatgtcaaccacaacttaactctgagctatcataa	2205
Db	84	CTCTGCTTAATTTGCTGCGCAGGATTTGTACTTCACITCCAACCTTACTCTGAGCTATTAATAA	25
Oy	2206	ataaacagatttatcttccagctt	2229
Db	24	ATAAACAGATTATTTTCCAGCTT	1
RESULT	13		
A1272081/c			
LOCUS	A1272081	385 bp	mRNA
DEFINITION	q17orf10.x1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1865611 3'	EST	17-MAR-1999
ACCESSION	A1272081		
VERSION	A1272081.1	GI:3891248	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 385) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
JOURNAL	On Jan 14, 1998 this sequence version replaced gi:11797673.		
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: www.bio.lml.gov/bbrp/lmage/lmage.html		
FEATURES	Insert Length: 568 Std Error: 0.00 Seg primer: -400P from GldpcO High quality sequence stop: 379. location/Qualifiers 1..385 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="IMAGE:1865611" /clone_id="NCI_CGAP_Kid3" /lab_host="DH10B" /note="Organ: kidney; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker; site_1: Not I; site_2: Eco RI; 1st strand cDNA was primed with a Not I...Oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."		
BASE COUNT	94 a	72 c	118 g
ORIGIN			101 t



Query Match 16.6%; Score 371.2; DB 44; Length 385;  
Best Local Similarity 97.9%; Pred. No. 9.2e-83;  
Matches 376; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1853 acatgacacacacagatgcatccttggacatgacatcaccagaatcacaac 1912  
DB 385 ACACACTACACACACAGATGCAATCCCTGGACCTGATATCAACCCAGGAATCTCAAC 326  
QY 1913 tcccttggccttgaaacagggccagataagaaacagctcgggaccccttttgaagggc 1972  
DB 325 TCCCTTGGCCCTGAAACAGGGCCAGATTAAGAAAGAGCTGGGCCCTCTTGAAGGCC 266  
QY 1973 aatgtgagaaagagagagcagccgcttgggaaagatcacaagatccagacttc 2032  
DB 265 AACGTGAGAGAAAGAGAGACGACCATTTGGAGAAATGTCAGATCCAGACTTC 206  
QY 2033 attccttctcttgcccaagatgaattgtctctccagcttggggagctccttcctg 2092  
DB 205 ATTCCCTTCTCTGGGCCAGTGAATTTGGTCTCTCCACGCTCGGGGAGCTCTCTTG 146  
QY 2093 aacctataaagaccccatgagatctctctctccatccctctcctctgacctgt 2152  
DB 145 AACCCATAAAGAACCCCACTGAGTCTCTCTCTCCATCCCTCTGCCCCCTGCT 86  
QY 2153 ctaattgtcagagatgtcaatccaacttactctgagctcataataaataaca 2212  
DB 85 CTAATTGCTGCGCAGATTTCTCACTCCAAACCTTACTCTGAGCTTAATAATAACA 26  
QY 2213 gattatttccagcttaaaaaa 2236  
DB 25 GATTATTCTTCCAGCTAAAAAAA 2

RESULT 14  
LOCUS AM390711/c 465 bp mRNA EST 04-FEB-2000  
DEFINITION RCI-ST0196-081099-011-cl1 ST0196 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM390711  
VERSION AM390711.1 GI:6895370  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 465)  
AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.  
TITLE The FAPESP/LICR Human Cancer Genome Project  
JOURNAL Unpublished (1999)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189704.  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC1-ST0196-081099-011-cl1&t3=1999-10-08&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 197  
High quality sequence stop: 464.  
Location/Qualifiers  
1. 465  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ST0196"  
/dev\_stage="Adult"  
/note="Organ: stomach; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products

## FEATURES

source

BASE COUNT 89 a 122 c 136 g 118 t  
ORIGIN

Query Match 16.6%; Score 371.2; DB 80; Length 465;  
Best Local Similarity 95.7%; Pred. No. 1e-82;  
Matches 446; Conservative 0; Mismatches 13; Indels 7; Gaps 6;

QY 1619 agatttgacacactacacacccctgctgagctgacgtcctctgacagagccctgaac 1678  
DB 465 AGATTGTGACACACTACACCTGCTGAGAGCTGACCTCCCTGCGAAGGCTGGAATC 406  
QY 1679 agggcgggactgag-alcgacctgtgtgtgcgcgtctgcccccaagagatagagat 1737  
DB 405 AGGGCGGGAGCTGGAATCGACCTTGTTGTCGCTGCCCCCAAGAGATGACGAT 346  
QY 1738 gagctgacctcttctgtgagtcacaaaggaggaacacatcagcgcttagaacaact 1797  
DB 345 GAGCTGACCTGCTT-CTGAAGTCCAAAAGGGGAACCAAAATTCAC-CGTTAGGAAACACT 288  
QY 1798 gagctcggccccaacctgtgacacaaagctcggaccagcttgagagagccacatg 1857  
DB 287 GAGCTCGGCCCCCACCCTCCTGTAACAAAGCCTGGATGACGCTTGAGAGGCCACT 228  
QY 1858 acacacacagatgcatcctctggagcctgaatcatalc-accaggaatcacaact-cc 1915  
DB 227 ACACACACACAGATGGGATCTTGGGACCTGAATCTCAACCCAGGAATCTCAACTCCC 168  
QY 1916 ctttgccctgaaacagggccagataaggaacagctcgagcc--acttttgaagggca 1973  
DB 167 CTTTGGCCCTGAACAGGGCCAGATTAAGGAACGACAGGGCCACTTCTTGAAGGCCA 108  
QY 1974 atgtgagaaagagagagcagccgcttgggaaagatcacaagatccagactctca 2033  
DB 107 AGGTGAGAGAAAGAGAGAGCCAGCCATTTGGAGAAAGTTCTCAAGATCCAGACTCTCA 48  
QY 2034 ttccttctcttgcccaagatgaattgtctctccagcttgggg 2079  
DB 47 TCCCTTCTCTGGGCCAGTGAATTTGGTCTCCACAGCTGAGGG 2

RESULT 15  
LOCUS A1453367/c 386 bp mRNA EST 09-MAR-1999  
DEFINITION t37cl1.x1 NCI\_CGAP\_Panl Homo sapiens cDNA clone IMAGE:2143700 3',  
mRNA sequence.  
ACCESSION A1453367  
VERSION A1453367.1 GI:4281343  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 386)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948675.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLN at:  
[www-bio.lnl.gov/db/brp/image/image.html](http://www-bio.lnl.gov/db/brp/image/image.html)



Seq primer: -40UP from Gibco  
High quality sequence stop: 338.  
Location/Qualifiers  
source  
1. 386

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2143700"  
/clone\_1lb="NCI CGAP Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: Salt;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 96 a 71 c 116 g 102 t 1 others  
ORIGIN

Query Match 16.5%; Score 369.2; DB 46; Length 386;  
Best Local Similarity 97.7%; Pred. No. 2.9e-82;  
Matches 374; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1854 catcacacacacacagatgcatccttggacatgacatcatalcaccaggaatcacaact 1913  
|||  
Db 386 CACTACACACACAGATGGCATCTTGGACCTGAATCTATCACCGAATCTCAAACT 327  
QY 1914 ccccttggcccttgaccagagccagataagaacagctcgagccactttttgaagcca 1973  
|||  
Db 326 CCCCTTGGCCCTTGACACAGGCGCAGATAAGGACAGCTCGGGCCACTTCTGAAGGCCA 267  
QY 1974 atgtggaggaaggaagcagcagccglttggaggaagatcgaagatcgaagctctca 2033  
|||  
Db 266 ACGTGAAGGAAGGAGGAGCAGCCACCATTTGGAGAAAGATCTCAAGATCCAGACTCTCA 207  
QY 2034 ttccttcctctggccagtgaaattgtctctctccagcttggggagactcctccttga 2093  
|||  
Db 206 TTCCTTCCCTGCGCCAGTGAATTGGTCTCTCCAGCTTGGGGAGCTCTCTTGA 147  
QY 2094 accctaataagaaccacacagctgagctctctctctccatccctctcctctgctc 2153  
|||  
Db 146 ACCCTAATAAGACCCCACTGAGTCTCTCTCTCCATCCCTCTCTGCTGCTT 87  
QY 2154 taattgtgcagagattgtcactccaaccttactctgagctcatataataataaag 2213  
|||  
Db 86 TAATTGCTGCCAGGATTGTCTCACTCAAACTTACTGTGAGCTCATTAATAATAACAG 27  
QY 2214 attatcttcoagcttaaaaaa 2236  
|||  
Db 26 ATTATTTTCCAGCTTAAAAAA 4

Search completed: May 19, 2000, 22:33:48  
Job time: 9649 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:27:48 ; Search time 158.18 Seconds  
(without alignments)  
2315.815 Million cell updates/sec

Title: US-09-502-945-2  
Perfect score: 2885  
Sequence: 1 ggaattccctctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCITUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.6	2.9	3489	US-08-728-323A-1	Sequence 1, Appl1
2	84.6	2.9	32207	US-08-770-379-20	Sequence 20, Appl1
3	79	2.7	5452	US-09-130-114-1	Sequence 1, Appl1
4	79	2.7	10596	US-07-884-811-15	Sequence 15, Appl1
5	79	2.7	10596	US-07-885-971-15	Sequence 15, Appl1
6	79	2.7	10596	US-08-087-783A-15	Sequence 15, Appl1
7	79	2.7	10596	US-08-194-088B-15	Sequence 15, Appl1
8	79	2.7	10596	US-08-194-087-15	Sequence 15, Appl1
9	79	2.7	10596	PCT-US93-04648-15	Sequence 15, Appl1
10	78.8	2.7	7218	US-08-232-463-14	Sequence 14, Appl1
11	72.6	2.5	1931	US-09-130-114-2	Sequence 1, Appl1
12	65	2.3	234	US-08-469-802B-3	Sequence 3, Appl1
13	65	2.3	234	US-08-267-803B-3	Sequence 3, Appl1
14	62	2.1	2338	US-08-425-069-1	Sequence 1, Appl1
15	62	2.1	2338	US-08-317-844B-1	Sequence 1, Appl1
16	60.4	2.1	2793	US-08-209-747-1	Sequence 1, Appl1
17	60.4	2.1	2793	US-08-458-298-1	Sequence 1, Appl1
18	60.2	2.1	543	5273901-6	Patent No. 5273901
19	59	2.0	1505	US-07-915-246-1	Sequence 1, Appl1
20	58.8	2.0	171	US-08-469-802B-5	Sequence 5, Appl1
21	58.8	2.0	171	US-08-267-803B-5	Sequence 5, Appl1
22	58.8	2.0	195	US-08-469-802B-2	Sequence 2, Appl1
23	58.8	2.0	195	US-08-267-803B-2	Sequence 2, Appl1
24	57	2.0	168	US-08-469-802B-4	Sequence 4, Appl1
25	57	2.0	168	US-08-267-803B-4	Sequence 4, Appl1
26	56.8	2.0	1995	US-08-425-069-3	Sequence 3, Appl1
27	56.8	2.0	1995	US-08-317-844B-3	Sequence 3, Appl1

28	56	1.9	9551	US-08-056-200-93	Sequence 93, Appl1
29	56	1.9	9551	US-08-800-644-93	Sequence 93, Appl1
30	55.4	1.9	2214	US-08-864-038A-1	Sequence 1, Appl1
31	55.4	1.9	3331	US-08-864-038A-2	Sequence 2, Appl1
32	55.4	1.9	3331	US-08-864-038A-4	Sequence 4, Appl1
33	55	1.9	1052	US-08-466-603-1	Sequence 1, Appl1
34	55	1.9	1052	US-08-314-503A-1	Sequence 1, Appl1
35	55	1.9	1052	US-08-468-066-1	Sequence 1, Appl1
36	55	1.9	1052	US-08-466-717-1	Sequence 1, Appl1
37	55	1.9	1052	PCT-US95-12414-1	Sequence 1, Appl1
38	54	1.9	336	US-07-814-220-4	Sequence 4, Appl1
39	54	1.9	336	US-07-812-421-4	Sequence 4, Appl1
40	54	1.9	533	5482709-5	Patent No. 5482709
41	53.6	1.9	12001	US-08-458-568A-11	Sequence 11, Appl1
42	52.8	1.8	2818	US-08-366-276-1	Sequence 1, Appl1
43	52.6	1.8	15378	US-08-785-420-1	Sequence 1, Appl1
44	52.4	1.8	336	US-07-814-220-3	Sequence 3, Appl1
45	52.4	1.8	336	US-07-812-421-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-08-728-323A-1  
; Sequence 1, Application US/08728323A  
; Patent No. 5948676  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
; TITLE OF INVENTION: Encoding Same And Uses Thereof  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA: US/08/728, 323A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: White, John P.  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-278-0400  
; TELEFAX: 212-391-0525  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3489 base pairs  
; TYPE: nucleic acid  
; TOPOLOGY: linear  
; STRANDEDNESS: single  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..3489  
; US-08-728-323A-1

	Query Match	2.9%	Score 84.6:	DB 4:	Length 3489:
	Best Local Similarity	48.8%:	Pred. No. 7.4e-11:		
	Matches 293:	Conservative 0:	Mismatches 299:	Indels 9:	Gaps 2
Qy	770	ctgcagacatgtgctgttctgtagcaggcccgccagcaagaccctcatlgtctgtcca	829		
Dd	1603	CAGCAGCAGGAGGCCACACAGCAGGAGGCACACAGCAGGAGCCACAGCAGGGGAGCCA	1662		
Qy	830	ctccagggcagltcccactagtcaacygttgaaagtlgtgccaccagatctcggacgta	889		
Dd	1663	CAGCAGCGGGAGCCCCACCACAGCGGGAGGCCACAGCAGCGGAGGCCA	1722		
Qy	890	ggcaagctcccgcgagcatcgcgccttgagcgcactagtcctaacgcgtgcgcag---	947		
Dd	1723	CAGCAGCGGGAGCCACACAGCGGGAGGCCACACAGCGGGAGGCCA	1782		
Qy	947	agtcceccagggcccttcagcagcttgttcatlycaacaacagcaccaagaftlcttgygaag	1006		
Dd	1783	CAGCAGCAGGATGAGCACAGCAGCAGGATGACGACGACGAGATGAGCAGCAGCATGAG	1842		
Qy	1007	cagaagcagcagcagctacagctggycaagatcctcaaccaagacagygagctgccccag	1066		
Dd	1843	CAGCAGCAGGATGACACAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCATGAG	1902		
Qy	1067	cagcccaaccaaccaacctgagagagcagaagagagctctcagagcagcagaaggtcttg	1126		
Dd	1903	CAGCAGCAGGATGACACAGCAGGATGAGCAGCAGGATGAGCAGCAGCATGAG	1962		
Qy	1127	ctgaggagagagagcccttgacatactgcccgaggaggtctccaacagagatagaagcacacag	1186		
Dd	1963	CAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCATGAG	2016		
Qy	1187	gaagaccctggagagagagacagagaaagagatggaggagagagattgtacatccag	1246		
Dd	2017	GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	2076		
Qy	1247	gttaagagccggagagagagaggtgtgtctgagaggggcccagacttgagagagcctggt	1306		
Dd	2077	GATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	2136		
Qy	1307	gctgatcaaaaactgttccagatgcccacacgcgtcacaactttgcaagtlgtacca	1366		
Dd	2137	GAGCAGCAGGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	2196		
Qy	1367	g 1367			
Dd	2197	G 2197			
<b>RESULT 2</b>					
US-08-770-379-20/c					
Sequence 20, Application US/0870379					
Patent No. 5849564					
<b>GENERAL INFORMATION:</b>					
APPLICANT: Chang, Yuan					
APPLICANT: Bohenzky, Roy A.					
APPLICANT: Russo, James J.					
APPLICANT: Edelman, Isidore S.					
APPLICANT: Moore, Patrick S.					
TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED					
TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF					
NUMBER OF SEQUENCES: 20					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: Cooper & Dunham LLP					
STREET: 1185 Avenue of the Americas					
CITY: New York					
STATE: New York					
COUNTRY: U.S.A.					
ZIP: 10036					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Floppy disk					
COMPUTER: IBM PC compatible					
OPERATING SYSTEM: PC-DOS/MS-DOS					

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.30
2
3 CURRENT APPLICATION DATA:
4 APPLICATION NUMBER: US/08/770,379
5
6 FILING DATE:
7 CLASSIFICATION: 435
8
9 ATTORNEY/AGENT INFORMATION:
10 NAME: White, John P.
11
12 REGISTRATION NUMBER: 28,678
13 REFERENCE/DOCKET NUMBER: 52342
14
15 TELECOMMUNICATION INFORMATION:
16 TELEPHONE: (212) 278-0400
17 TELEFAX: (212) 391-0525
18
19 INFORMATION FOR SEQ ID NO: 20:
20
21 SEQUENCE CHARACTERISTICS:
22
23 LENGTH: 32207 base pairs
24 TYPE: nucleic acid
25 STRANDEDNESS: double
26
27 TOPOLOGY: linear
28
29 MOLECULE TYPE: DNA (genomic)
30
31 US-08-770-379-20

```

Query Match	Similarity	2.9%	Score 84.6	DB 3	Length 32207
Best Local	48.8%	Pred. No. 1.8e-10			
Matches 293	Conservative	0	Mismatches 299	Indels	9
					Gaps 2
QY 770	ctcgcagcatgtgtctgtcgtgagcagggcccgccagcagcagagaccctcattgtctgtgcca	829			
Db 20394	CAGCAGCAGAGGAGCCACACAGCAGCAGGAGGCCACACAGCAGCAGGAGCCACACAGCAGCGGAGCCA	20335			
QY 830	ctccacgggacagttcccacatagtgaacggtgaagctgttgccaccagcatgtggagcgtga	889			
Db 20334	CAGCAGCGGGAGCCCCACACAGCGGGAGCCACAGCGGGAGCCACACAGCAGCGGGAGCCA	20275			
QY 890	ggcaagctcccccgggcacatcgcgcctctgagccgcacatcaagttcctaccgctgcgcgca---	947			
Db 20274	CAGCAGCGGGAGCCACACAGCAGCGGGAGCCACACAGCAGCGGGAGCCACACAGCAGCGGGAGCCA	20215			
QY 947	agtcgccagggcccttcagcagcgtgtgtcaltgcacaacacagcaccagcagttctcttgagaag	1006			
Db 20214	CAGCAGCAGGAGTGAACAGCAGCAGGAGTGAAGCAGCAGCAGGAGTGAAGCAGCAGGAGTGAAG	20155			
QY 1007	cagaagcagcagcagctacagcttgggcaagatcctccaccagaacagggagctgtgccag	1066			
Db 20154	CAGCAGCAGGAGTGAACAGCAGCAGGAGTGAAGCAGCAGCAGGAGTGAACAGCAGGAGTGAAG	20095			
QY 1067	cagccaccaccaccaccctctgagagagacagagggagctcagcagcagcagggagctctgt	1126			
Db 20094	CAGCAGCAGGAGTGAACAGCAGCAGGAGTGAAGCAGCAGCAGGAGTGAAGCAGCAGGAGTGAAG	20035			
QY 1127	ctgggggagagagagcccttgaccatctgcgccgggaggggtccaccagagagtgagagacacag	1186			
Db 20034	CAGCAGCAGGAGTGAAC-----AGCAGCAGGAGTGAAGCAGCAGGAGTGAAGCAGCAGCAG	1981			
QY 1187	gaagaccctggagagagagagacgaggaagaagaagtgtggagagagagagagatttcacccag	1246			
Db 19980	GATGAGCAGCAGCAGCAGGAGTGAAGCAGCAGCAGGAGTGAAGCAGCAGGAGTGAAGCAGCAG	19921			
QY 1247	gttaaggacagagagagggcgagagtggtgtctgagagagggggcccgaccttgagagagccttgt	1306			
Db 19920	GATGAGCAGCAGCAGGAGTGAAGCAGCAGCAGGAGTGAAGCAGCAGGAGCAGCAGGAGTGAAGCAG	19861			
QY 1307	gctgatatcaaaaactgttctcagatgtcccacacgctctgcaaccttctgcaggtgtaccaa	1366			
Db 19860	GAGCAGCAGGAGTGAACAGCAGCAGGAGTGAAGCAGCAGCAGGAGTGAAGCAGCAGCAG	19801			
QY 1367	g 1367				
Db 19800	G 19800				







Query Match	2.7%	Score 79;	DB 6;	Length 10536;
Best Local Similarly	51.3%	Pred. No. 2.4e+09;		
Matches 184; Conservative	0;	Mismatches 175;	Indels 0;	Gaps 0;

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match 2.7%, Score 78.8; DB 1; Length 7218;
Best Local Similarity 3.6%, Pred.No.2,3e-09;
Matches 14; Conservative 241; Mismatches 133; Indels 0; Gaps 0;

Oy 963 agcagctgtgtcgtcaacaacagccagcagcttcctgagagcagcagcagcagc 1022
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1428 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1369

Oy 1023 tacagctgtgcaagatcctcaccaagagagagctgtccagcagccacccacc 1082
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1368 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1309

Oy 1083 ctgagagagacagagagagctgtcacgagcagcagagagcttctctgtggagagcc 1142
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1308 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1249

Oy 1143 tgacatgccccggagggtccacagagagtggagcacaaggaagacctgtgagag 1202
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1248 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1189

Oy 1203 aggaagagaaagagatgtgggagagagagatgtcaccaggttaagcagagag 1262
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1188 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1129

Oy 1263 gcgagagtggtgtctgaggaaggccgacctgtgagagagctgtgtgatacaaaaac 1322
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1128 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1069

Oy 1323 tgttcagatgtcccaacgcgtgcagcc 1350
   :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
Db 1068 RRATCGAAGCTCCTCGACTGCAGCC 1041

RESULT 11
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes

```



```

? TITLE OF INVENTION: From Multiple Transfected Episomes
?
? FILE REFERENCE: 0867/1D9030U1
?
? CURRENT APPLICATION NUMBER: US/09/130,114
?
? CURRENT FILING DATE: 1998-08-06
?
? NUMBER OF SEQ ID NOS: 36
?
? SOFTWARE: FastSeq for Windows Version 3.0
?
? SEQ ID NO 2
?
? LENGTH: 1931
?
? TYPE: DNA
?
? ORGANISM: EBNA
?
? US-09-130-114-2

```

Query Match	2.5%;	Score 72.6;	DB 4;	Length 1931;
Best Local Similarity	48.6%;	Pred. No. 3.8e-08;		
Matches 198; Conservative	0;	Mismatches 209;	Indels 0;	Gaps 0;

[illegible]

RESULT 12  
 US-08-469-802B-3  
 Sequence 3, Application US/08469802B  
 Patent No. 5741645  
 GENERAL INFORMATION:  
 APPLICANT: Orr, Harry T.  
 APPLICANT: Rannu, Laura P.W.  
 APPLICANT: Chung, Ming-Yi  
 APPLICANT: Zoghbi, Huda Y.  
 TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
 Patent No. 5741645  
 TITLE OF INVENTION: Type 1 and Method for Diagnosis  
 NUMBER OF SEQUENCES: 47  
 CORRESPONDENCE ADDRESS:  
 ADDRESSSEE: Muelting, Raasch, Gebhardt & Schnappach, P.A.  
 STREET: 119 No. 5741645th Fourth Street, Suite 203  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55401  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/469, 802B

1 FILING DATE: 06-JUN-1995  
 2 CLASSIFICATION: 435  
 3 ATTORNEY/AGENT INFORMATION:  
 4 NAME: Mueeling, Ann M.  
 5 REGISTRATION NUMBER: 33,977  
 6 REFERENCE/DOCKET NUMBER: 110.000301071  
 7 TELECOMMUNICATION INFORMATION:  
 8 TELEPHONE: 612-305-1217  
 9 TELEFAX: 612-305-1225  
 10 INFORMATION FOR SEQ ID NO: 3:  
 11 SEQUENCE CHARACTERISTICS:  
 12 LENGTH: 234 base pairs  
 13 TYPE: nucleic acid  
 14 STRANDEDNESS: single  
 15 TOPOLOGY: linear  
 16 MOLECULE TYPE: DNA  
 17  
 18 US-08-469-802b-3

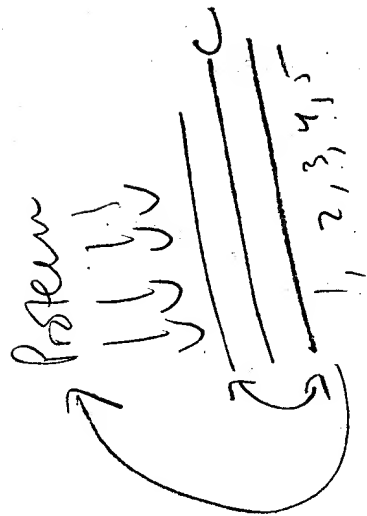
Query Match	2.3%;	Score 65;	DB 2;	Length 234;
Best Local Similarity	55.6%;	Pred. No. 1e-06;		
Matches 125;	Conservative	0;	Mismatches 100;	Indels 0;
			Gaps	0;

[illegible]

US-RESULT 13  
; Sequence 3, Application US/08267803B  
; Patent No. 5834183  
; GENERAL INFORMATION:  
; APPLICANT: Orr, Harry T.  
; APPLICANT: Ranum, Laura P.W.  
; APPLICANT: Chung, Ming-yi  
; APPLICANT: Zoghbi, Huda Y.  
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia  
; Patent No. 5834183  
; TITLE OF INVENTION: Type 1 and Method for Diagnosis  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Mueling, Raasch, Gebhardt & Schwappach, P.A.  
; STREET: P.O. Box 581415  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55458-1415  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/267, 803B  
; FILING DATE: 28-JUN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: McCormack, Myra H.  
; REGISTRATION NUMBER: 36,602  
; REFERENCE/DOCKET NUMBER: 110.00030120







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 19:52:59 ; Search time 3256.04 Seconds  
(without alignments)  
1931.980 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 1552  
Sequence: 1 cttctggatgcacgcgagaa.....aaatgaacttttaagaaa 1552

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: gb\_est1:\*  
21: gb\_est2:\*  
22: gb\_est3:\*  
23: gb\_est4:\*  
24: gb\_est5:\*  
25: gb\_est6:\*  
26: gb\_est7:\*  
27: gb\_est8:\*  
28: gb\_est9:\*  
29: gb\_est10:\*  
30: gb\_est11:\*  
31: gb\_est12:\*  
32: gb\_est13:\*  
33: gb\_est14:\*  
34: gb\_est15:\*  
35: gb\_est16:\*  
36: gb\_est17:\*  
37: gb\_est18:\*  
38: gb\_est19:\*  
39: gb\_est20:\*  
40: gb\_est21:\*  
41: gb\_est22:\*  
42: gb\_est23:\*  
43: gb\_est24:\*  
44: gb\_est25:\*

45: gb\_est26:\*  
46: gb\_est27:\*  
47: gb\_est28:\*  
48: gb\_est29:\*  
49: gb\_est30:\*  
50: gb\_est31:\*  
51: gb\_est32:\*  
52: em\_est20:\*  
53: em\_est21:\*  
54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*  
59: gb\_est33:\*  
60: gb\_est34:\*  
61: gb\_est35:\*  
62: gb\_est36:\*  
63: gb\_est37:\*  
64: gb\_est38:\*  
65: em\_est27:\*  
66: em\_est28:\*  
67: em\_est29:\*  
68: em\_est30:\*  
69: gb\_est39:\*  
70: gb\_est40:\*  
71: gb\_est41:\*  
72: gb\_est42:\*  
73: gb\_est43:\*  
74: gb\_est44:\*  
75: em\_est31:\*  
76: em\_est32:\*  
77: em\_est33:\*  
78: em\_est34:\*  
79: gb\_est45:\*  
80: gb\_est46:\*  
81: gb\_est47:\*  
82: gb\_gss1:\*  
83: gb\_gss2:\*  
84: gb\_gss3:\*  
85: gb\_gss4:\*  
86: em\_gss1:\*  
87: em\_gss2:\*  
88: em\_gss3:\*  
89: em\_gss4:\*  
90: gb\_gss5:\*  
91: gb\_gss6:\*  
92: gb\_gss7:\*  
93: gb\_gss8:\*  
94: gb\_gss9:\*  
95: em\_gss5:\*  
96: em\_gss6:\*  
97: em\_gss7:\*  
98: em\_gss8:\*  
99: em\_gss9:\*  
100: em\_gss10:\*  
101: em\_gss11:\*  
102: gb\_gss10:\*  
103: gb\_gss11:\*  
104: em\_gss12:\*  
105: gb\_gss12:\*  
106: gb\_gss13:\*  
107: gb\_gss14:\*  
108: gb\_gss15:\*  
109: gb\_gss16:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result

Query

No.	Score	Match	Length	DB	ID	Description
C 1	465.2	30.0	490	51	A1735499	A1735499 a112006.x.s1
C 2	412	26.5	419	64	AM028371	AM028371 wv8b509.x.s1
C 3	371.4	23.9	410	46	A1445418	A1445418 t_j34b31.t
C 4	369	23.8	391	42	A1093201	A1093201 qab0a09.x.s1
C 5	349.4	22.3	365	44	A1266387	A1266387 qm45h08.x.s1
C 6	345.4	22.3	360	44	A1888695	A1888695 q181c08.x.s1
C 7	335.8	21.6	398	39	AA886109	AA886109 y474f05.s1
C 8	328.4	21.2	433	42	A1140814	A1140814 ooi15c05.x.s1
C 9	327.2	21.1	384	41	A1038590	A1038590 ox34e03.s1
C 10	326.4	21.0	424	39	AA828894	AA828894 oes1e12.s1
C 11	326.4	21.0	430	39	AA879456	AA879456 o_j91p03.s1
C 12	326.4	21.0	439	27	AA029201	AA029201 z_k12f08.s1
C 13	326.4	21.0	455	28	AA1023109	AA1023109 z_k87g11.s1
C 14	326.4	21.0	462	43	A1168222	A1168222 ok30c01.x.s1
C 15	326.4	21.0	472	25	NS1485	NS1485 yz04e06.s1
C 16	326.4	21.0	482	42	A1148323	A1148323 qd31d07.x.s1
C 17	326.4	21.0	487	39	AA843811	AA843811 ak09c08.s1
C 18	326.4	21.0	487	44	A1298647	A1298647 qm90c10.x.s1
C 19	326.4	21.0	569	42	A1089332	A1089332 qb16g08.x.s1
C 20	326.4	21.0	596	26	w72147	w72147 zd70f08.s1
C 21	325.4	21.0	587	74	AM193663	AM193663 km29p02.x.s1
C 22	324.8	20.9	418	60	A1808317	A1808317 w_f54d08.x.s1
C 23	324.8	20.9	442	41	A1033069	A1033069 ow93f02.s1
C 24	323.8	20.9	499	25	NS1277	NS1277 yz14d07.s1
C 25	323	20.8	482	74	AM195719	AM195719 km85b07.x.s1
C 26	321.8	20.7	556	29	AA161645	AA161645 zc73a06.s1
C 27	321.6	20.7	437	38	AA744290	AA744290 nv51d02.x.s1
C 28	320.6	20.7	572	63	A1970343	A1970343 w91a04.x.s1
C 29	317.4	20.5	449	22	R71679	R71679 yj85e08.s1
C 30	317.4	20.5	444	38	AA813519	AA813519 a179p03.s1
C 31	315.4	20.3	422	25	NC4578	NC4578 yz91e06.s1
C 32	315.4	20.3	436	24	H99906	H99906 yx32h10.s1
C 33	313.4	20.2	472	28	AA095913	AA095913 z_k87g11.t
C 34	311.4	20.1	451	64	AM021346	AM021346 dt21c04.y.s1
C 35	308.2	19.9	345	39	AA883684	AA883684 a158a05.s1
C 36	304.4	19.6	377	44	A1267979	A1267979 g044f10.x.s1
C 37	304.4	19.6	391	28	AA083859	AA083859 zm16d06.s1
C 38	296.6	19.1	407	22	R36854	R36854 yf52c07.s1
C 39	292.8	18.9	451	28	AA089195	AA089195 mc05h11.t
C 40	292.2	18.8	432	22	R39448	R39448 yc95d03.s1
C 41	288.2	18.6	372	32	AA346369	AA346369 est52776
C 42	287.2	18.5	468	24	H98664	H98664 yx17g01.s1
C 43	284.4	18.3	322	44	AA910762	AA910762 o125p06.s1
C 44	278.4	18.0	465	21	R07471	R07471 ye97a06.s1
C 45	278.4	17.9	295	49	F25334	F25334 hspd12318.h

## ALIGNMENTS

	<p>Contact: Wilson RK  Washington University School of Medicine  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  Tel: 314 286 1800  Fax: 314 286 1810  Email: est@watson.wustl.edu  This clone is available royalty-free through LNLN ; contact the  IMAGE Consortium (info@image.llnl.gov) for further information.  Possible reversed clone: similarity on wrong strand  Seq primer: -40UP from Gibco  High quality sequence stop: 344.</p>
FEATURES	
Source	<p>Location/Qualifiers</p> <p>1..490</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="IMAGE:2354891"</p> <p>/clone_lib="Barstead aorta HPLRB6"</p> <p>/sex="male"</p> <p>/sex-stage="adult, age 64"</p> <p>/lab_host="DH10B (phage resistant)"</p> <p>/note="Organ: aorta; Vector: pT73D-Pac (Pharmacia) with a  modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st  strand cDNA was primed with a Not I - oligo(dT) primer [5'  TGTTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTTCTTTTCTTTT  3'], double-stranded cDNA was ligated to Eco RI adaptors  5' AATTGGATGGAAC 3' and 5' GTTGGATCGG 3'], digested  with Not I and cloned into the Not I and Eco RI sites of  the modified pT73 vector. Library constructed by Bob  Barstead."</p>
BASE COUNT	111 a 130 c 119 g 130 t
ORIGIN	
Query Match	30.0%; Score 465.2; DB 51; Length 490;
Best Local Similarity	98.2%; Pred. No. 5.5e-110;
Matches 481; Conservative	0; Mismatches 8; Indels 1; Gaps 1;
QY 1060	gggaaagatcatatgaacagatgaagaaaggtctaaagcagcttggatgaacagccagggc 1119
Db 490	gggaaagatcatatgaacagatgaagaaaggtctaaagcagcttggatgaacagccagggc 431
QY 1120	acagccagcagctgtgtcagctctctcagcaagcagaacacagctctctctgtgaagagcag 1179
Db 430	ACAGCCGACGACGCTGGTGCAGCTCTTCAGCAAGCAAGCAACGACTTCTCTGTGAGAGGCGAG 371
QY 1180	agctctgtcggaaagagtgtagccgctctgcggaccagttaccagcagtcacacatctgat 1239

[illegible]

Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LUNL; contact the  
 IMAGE Consortium (infoimage.lunl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -400P from Glibco  
 High quality sequence stop: 344.  
 Location/Qualifiers  
 1. 490  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2354891"  
 /clone\_1lb="Barstead aorta HPLRB6"  
 /sex="male"  
 /dex\_stage="adult, age 64"  
 /lab\_host="DH10B (phage resistant)"  
 /note="Organ: aorta; Vector: pT73D-Pac (Pharmacia) with a  
 modified polylinker; Site\_1: EcoRI; Site\_2: NotI; 1st  
 strand cDNA was primed with a Not I - oligo(dT) primer [5'  
 TGTACGATCATCTGAAGTGGGAGCGCGCCCTTTTCTTTTCTTTTCTTTT  
 3']; double-stranded cDNA was ligated to Eco RI adaptors  
 [5' AATTCGATCGAAC 3' and 5' GTTCCATCGG 3'], digested  
 with Not I and cloned into the Not I and Eco RI sites of  
 the modified pT73 vector. Library constructed by Bob  
 Barstead."

BASE COUNT	111 a	130 c	119 g	130 t
ORIGIN				
Query Match	30.0%;	Score 465.2;	DB 51;	Length 490;
Best Local Similarity	98.2%;	Pred. No. 5.5e-110;		
Matches 481;	Conservative 0;	Mismatches 8;	Indels 1;	Gaps 1;
Db 1060	gggaagatcatatgagacgatgaagcaaaagctgaagcagctgataaagcacagccagcc	1119		
Db 490	gggaagatcatatgagacgatgaagcaaaagctgaagcagctgataaagcacagccagcc	431		
Qy 1120	acagaccagcagctgtgtcagctctctcaagcagcaagcaccagcttctctgaagagcgag	1179		
Db 430	acagcccgacacgctgtgtcagctctctcaagcagcaagcaccagcttctctgaagagcgag	371		
Qy 1180	agccgtctggaagaggtgagaccggtctggaagccagttaccagcatgccaacatctgat	1239		
Db 370	agccgtctggaagaggtgagaccggtctggaagccagttaccagcatgccaacatctgat	311		
Qy 1240	tcgtcaccttgatggaacagagtgaaataatgaaattacaagaagatatattacattc	1299		
Db 310	tcgtcaccttgatggaacagagtgaaataatgaaattacaagaagatatattacattc	251		
Qy 1300	tggtttaactaataatgaccacaagcaccagaccttccagaggtgacacccgctcagc	1359		
Db 250	tggtttaactaataatgaccacaagcaccagaccttccagaggtgacacccgctcagc	191		
Qy 1360	ctgcagatgagcctgtgtctctcaacagcgagcgctgtctcccgcaagcagctcgagctgag	1419		
Db 190	ctgcagatgagcctgtgtctctcaacagcgagcgctgtctcccgcaagcagctcgagctgag	131		
Qy 1420	ctggaagctcagctcagctgaagcaga-ctcctgtgtatgtttttcagaagaatgcttgaag	1478		
Db 130	ctggaagctcagctcagctgaagcaga-ctcctgtgtatgtttttcagaagaatgcttgaag	71		
Qy 1479	ttatgtgtttaaattctgtctcatttgtatgtcaggttatacatatgattttcaataaaga	1538		
Db 70	ttatgtgtttaaattctgtctcatttgtatgtcaggttatacatatgattttcaataaaga	11		
Qy 1539	actttttaa 1548			
Db 10	actttttaa 1			

RESULT 2  
AM028371/c  
LOCUS AM028371 479 bp mRNA EST 15-SEP-1999  
DEFINITION w988b09.x1 NCI\_CGAP\_Gas4 Homo sapiens cDNA clone IMAGE:2541209 3'  
Similar to TR:060527 060527 ANTIGEN NY-CO-8 ;, mRNA sequence.  
ACCESSION AM028371  
VERSION AM028371.1 GI:5887127  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 479)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 22, 1998 this sequence version replaced gi:3247399.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:  
[www-bio.illn1.gov/bbrp/image/image.html](http://www-bio.illn1.gov/bbrp/image/image.html)

FEATURES  
source  
Seg primer: -40UP from Gibco  
High quality sequence stop: 361.  
Location/Qualifiers  
1..479  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2541209"  
/clone\_lib="NCI\_CGAP\_Gas4"  
/tissue\_type="poorly differentiated adenocarcinoma with  
signed ring cell features"  
/lab\_host="DH10B"  
/note="Organ: stomach; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.69 kb. Life Technologies catalog #:  
11549-011"

BASE COUNT 113 a 121 c 118 g 127 t  
ORIGIN

Query Match 26.5%; Score 412; DB 64; Length 479;  
Best Local Similarity 95.0%; Pred. No. 3.3e-96;  
Matches 458; Conservative 0; Mismatches 20; Indels 4; Gaps 3;

QY 1069 catgagagatgaagcaagctgaagcagctggtgaagcagcagcagccag 1128  
|||||  
Db 479 CATGAGAGATGAACCAAGCTTAAAGCAGCTTGATTA--GCCCAAGAGGACACAGCCAGC 422  
QY 1129 cagctggtgagctcctcctaagcagcagcagctctcctctgagagcagcagctgtcg 1188  
|||||  
Db 421 AAGCTGGTGCAGCTCTCTGACGACGAAACGAGCTTCTCTGAGAGGACAGAGCCT-TCG 363  
QY 1189 gaagaggtggaacggctcgcgaacccagttaccagcagtcgcaaatctgattgtgaact 1248  
|||||  
Db 362 GAAGAGGTGAGACCGGCTCGCGACCCAGTTACCCAGCATGCCACATCTGATGTGTGACCT 303  
QY 1249 ggaagagcaagatgaataatgaatcacaagaagatataatcattatcctggtttga 1308  
|||||  
Db 302 GGATGGAACAGAGTGAATTAATGATTTACAAAGAGATTTTACATTCGTGGTTTGA 243  
QY 1309 cttaaatgacacagcagcagcagcctcccaagggtgacacagcagcagcctgagtgag 1368  
|||||  
Db 242 CTTATATATGACACAGCAGACGACTTCCAGGAGTGAACCGGCTCAGCCTGCAAGTGG 183

QY 1369 ggcctgtcctcatcaacagcggcgctgtcccccagcagcagtcggcgtgagctgagct 1428  
|||||  
Db 182 GGCTGTCTCTCAACACGCGGGCTGTCCCGCAGCAGAGTGGGCTGGAGCTGAAGCT 123  
QY 1429 gactcagctgagcaga-ctccctgtgtatgttttcagagaatgagcttaagttatgtt 1487  
|||||  
Db 122 GACTCTAGCTGAGCAGAGCTCTCTGCTATGTGTTTCAGAAATGCTTGAAGTTATGTGT 63  
QY 1488 taatctgtcattcgtatgctaggttatcatatgatattcaataatgaaacttttaa 1547  
|||||  
Db 62 TAAACTGCTCATTTGGTATGCTAGGTTATACATATGATTTTCAATAAATGAACCTTTTAA 3  
QY 1548 ag 1549  
||  
Db 2 AG 1

RESULT 3  
A1445418/c  
LOCUS A1445418 410 bp mRNA EST 13-APR-1999  
DEFINITION t3j4b11.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2143389 3',  
mRNA sequence.  
ACCESSION A1445418  
VERSION A1445418.1 GI:4288073  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 410)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 7, 1998 this sequence version replaced gi:3121415.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Life Technologies catalog #: 11548-013  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:  
[www-bio.illn1.gov/bbrp/image/image.html](http://www-bio.illn1.gov/bbrp/image/image.html)

FEATURES  
source  
Insert Length: 2579 Std Error: 0.00  
Seg primer: -40UP from Gibco  
High quality sequence stop: 368.  
Location/Qualifiers  
1..410  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2143389"  
/clone\_lib="NCI\_CGAP\_Pan1"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Average insert size 1.72 kb. Life Technologies catalog #:  
11548-013"

BASE COUNT 100 a 99 c 91 g 120 t  
ORIGIN

Query Match 23.9%; Score 371.4; DB 46; Length 410;  
Best Local Similarity 99.5%; Pred. No. 1e-85;  
Matches 383; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1166 tccctgagagcagagcctgtctgcgaagaggtgagccggtcgcgacccagttaccacga 1225  
|||||  
Db 410 TCCTGGAAGGAGGAGCGCTGTCGGAAGAGTGAACCGGCTCGGACCAAGTTACCCACGA 351  
QY 1226 tgcacaatctgattgtcgtgacctgtgagtggaacagagtgtaataatgataacaaagaga 1285

```
|||||
Db 350 TGCCAACTGATGCTGACCTGGATGGAACAGAGTAATAATGATTTACAAAGAGA 291
Qy 1286 tatttacctatctgtgttgacttaataatgccaacagcagccctcccaaggt 1345
Db 290 TATTTTACATTCATCGTGTGACTTAAATGACCAACAGCAGCAGCTTCCAGAGGT 231
Qy 1346 gaacccgctcagccctcagtgaggctgtctctcacaacgcgagcgtccccagc 1405
Db 230 GACACCGCTCAGCCTGCAAGTGGGGCTGTCTCTCATCAACCGCGGCTGTCCCGCAGC 171
Qy 1406 cagtcggagctgagcgtgagctgactctagctgagcaga-ctccgtgtatgtttca 1464
Db 170 CAGTCGGCTGAGCTGAGCTGACTGACTGAGCTGAGCAGAGCTCCTGTGATGTTTCA 111
Qy 1465 gaattgcttaagttatgttttaaatctctatctgtatgtcaggttatcataga 1524
Db 110 GAATGCGCTTGAAGTATGTTTAAATCTGCTCATTTGCTAGGTATACATATGA 51
Qy 1525 ttccaataatgaacttttaag 1549
Db 50 TTTTCATTAATGAACITTTTAAAG 26

RESULT 4
LOCUS AI092201 391 bp mRNA EST 05-OCT-1998
DEFINITION ga60a09.s1 Soares_fetal_heart_NbH19W Homo sapiens cDNA clone
IMAGE:1691128 3, mRNA sequence.
ACCESSION AI092201
VERSION AI092201.1 GI:3431195
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 391)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045437.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 503 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 297.
Location/Qualifiers
1.391
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1691128"
/clone_1bp="Soares_fetal_heart_NbH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATCTGAGTGAAGGAGCGCGCATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbH19W."
BASE COUNT 87 a 102 c 65 g 137 t
ORIGIN
```

```
Query Match 23.8%; Score 369; DB 42; Length 391;
Best Local Similarity 100.0%; Pred. No. 4,3e-85;
Matches 369; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 172 cgtgtgtgtgtcttcttltgaatgtgtcagcatgaagctgttcttcccaaccat 231
Db 391 CGTGTGTGTGTCCTTTGTTGAATGTGCTCAGCATGAAGCTGTTCTTCCCAACCAT 332
Qy 232 actaatgtcatatgcagacacatgcgaagactggtttaagaagaatgactgtatct 291
Db 331 ACTAATGTTCAATGACAGACCATCGAAAGACTGTTAAAGAAAGATGACTGATGCT 272
Qy 292 gcaatgtttccgtaagaagagactgtgcagatgcgaagcaagaagaagagctat 351
Db 271 GCATGAGTTTCCGTAAAGAGAGCTGTGGCAATACGACAGCAAGAGAGAGTGTAT 212
Qy 352 gaacagtgaaacaagttttcaatatctgaggaaagccaatttgaaaaaccaagct 411
Db 211 GACAGGTGAACAAAGTTTTCATAATCTGAGGAGACCATTTTGAAAAACCAAGCT 152
Qy 412 ttaatcagtgtagccagtgtaggaagagcgtgagagcgagcgacttgaataa 471
Db 151 TTAATCCAGTGTGACACGTTTGAGGAGAGAGCTGAGAGCGCAGCGACTTGAAAA 92
Qy 472 gaattgcatctcagcaagagaagagccattgaagaagacatgataaagaata 531
Db 91 GAAGCTTGATCTCAGCAAGAGAAAAAGGCCATTGAGAAAGCATGATGAAAAAGAAATA 32
Qy 532 accgaagaa 540
Db 31 ACGAAAGAA 23

RESULT 5
LOCUS AI266387/c 365 bp mRNA EST 29-JAN-1999
DEFINITION gm45h08.x1 Soares_placenta_8to9weeks_2NbH8to9W Homo sapiens cDNA
IMAGE:1891743 3, mRNA sequence.
ACCESSION AI266387
VERSION AI266387.1 GI:3884545
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 365)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2282340.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1794 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 361.
Location/Qualifiers
1.365
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1891743"
/clone_1bp="Soares_placenta_8to9weeks_2NbH8to9W"
/dev_stage="two placenta: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
```



TTTACCAATCTGAAGTGGAGCGCCGCAATTTTTTTTTTTTTT 3',  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M.Fatima Bonaldo."

BASE COUNT 96 a 89 c 83 g 97 t  
ORIGIN

Query Match 22.5%; Score 349.4; DB 44; Length 365;  
Best Local Similarity 99.4%; Pred. No. 5e-80;  
Matches 361; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1191 agaggtgacggcgtcgccgagccagttaccagcatgccacaattgattctgacccgg 1250  
|||||  
Db 365 AGAGGTGACCGCGTCGCGACCCAGTACCCGCAATGCGCATCTGATTCGTGACCTGG 306  
Oy 1251 atggacagagtgaaataatgaataaagaatattatcatctgtgttgaact 1310  
|||||  
Db 305 ATGGACAGAGTGAATTAATGATTACAAAGATATTATTCATCTGCTTTAGACT 246  
Oy 1311 taatgtccacaacgacacagacccctcccaagggtgacacgcgctcagcctgagtgagg 1370  
|||||  
Db 245 TAAATATGCCAACACGACACGACCTTCCAGAGGTGACACGCGCTCAGCTGAGGGG 186  
Oy 1371 ctgtctccatcaacggcgaggcgtctcccgacgacgctgagctgagctgagctcga 1430  
|||||  
Db 185 CTGGTCCATCAATCAACGCGGCGCTCTCCCGCACGACGCGGGGAGCTGAGCTGGA 126  
Oy 1431 ctctagctgagcaga-ctcctgtgtatgttttcagaagaatgctgaaagttaagtta 1489  
|||||  
Db 125 CTCTAGCTGAGCAGACGCTCTGCTGCTGTTTTCAGAAATGCTTGAAGTTATGTTTA 66  
Oy 1490 aatctgctcatctgtagtctaggtatatacatatgatttcaataaagaacttttaag 1549  
|||||  
Db 65 AATCGTCATTCGTATGTTAGTTATACATATGATTTTCATTAATGAACTTTTAAAG 6  
Oy 1550 aaa 1552  
|||||  
Db 5 AAA 3

RESULT 6  
AI288955/c 360 bp mRNA EST 29-JAN-1999  
LOCUS q181c08.x1 Soares\_NHMPu\_S1 Homo sapiens cDNA clone IMAGE:1878734  
DEFINITION 3', mRNA sequence.  
ACCESSION AI288955  
VERSION AI288955.1 GI:3931264  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 360)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Feb 11, 1998 this sequence version replaced gi:2873143.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Insert Length: 431 Std Error: 0.00  
Seq primer: -40UP from Gluco.

FEATURES  
source  
1. .360  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1878734"

/clone\_11b="Soares\_NHMPu\_S1"  
/tissue\_type="Pooled human melanocyte, fetal heart, and  
pregnant uterus"  
/lab\_host="DH10B"  
/note="Organ: mixed (see below); Vector: pT7T3D-Pac  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; Equal amounts of plasmid DNA from three  
normalized libraries (melanocyte 2NbH, pregnant uterus  
NBHPU, and fetal heart NBHRIW) were mixed, and ss circles  
were made in vitro. Following HAP purification, this DNA  
was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from pools of  
5,000 clones made from the same 3 libraries. The pools  
consisted of T.M.A.G.E. clones 260232-265223,  
340488-345479, and 484488-489479."

BASE COUNT 96 a 88 c 84 g 92 t  
ORIGIN

Query Match 22.3%; Score 345.4; DB 44; Length 360;  
Best Local Similarity 99.4%; Pred. No. 5.4e-79;  
Matches 357; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Oy 1193 aggtgacggcgtcgccgagccagttaccagcatgccacaatctgattgctgactgag 1252  
|||||  
Db 360 AGGTGACCGCGCTCGCGACCCAGTACCCAGCATGCCAATCTGATTCGACTGGAT 301  
Oy 1253 ggaacagagtgaataatgaataaagaatattacattcctgtgttaactta 1312  
|||||  
Db 300 GGAACAGAGTGAATTAATGATTTCACAAAGAGATTTTATCATCTACCTGTTAGACTTA 241  
Oy 1313 atatgccacaacgacacagacccctcccaagggtgacacgcgctcagcctgagtgagg 1372  
|||||  
Db 240 AATATGCCAACACGACACGACCTTCCAGGGTGAACACGCTCAGCTGAGCTGGGCT 181  
Oy 1373 ggtctctcaacaggcgaggcgtctcccgacgacgctgagctgagctgagcttga 1432  
|||||  
Db 180 GGTCTCATCAACGCGGCGCTGCTCCCGCACGCTCGGCTGGAGCTGAGCTGACT 121  
Oy 1433 ctagctgagcaga-ctcctgtgtatgttttcagaagaatgctgaaagttaagtta 1491  
|||||  
Db 120 CTAGCTGAGCAGACGCTCGGTGTTTTCAGAAATGCTTGAAGTTATGTTTAA 61  
Oy 1492 tctgtcattgtatgctaggttatacatatgatttcaataaagaacttttaaga 1550  
|||||  
Db 60 TCTGCTCATTCGTATGTTAGTTATACATATGATTTTCATTAATGAACTTTTAAAGA 2

RESULT 7  
AA886109 398 bp mRNA EST 30-MAR-1998  
LOCUS ny44f05.s1 NCI-CCAP\_Pri12 Homo sapiens cDNA clone IMAGE:1274625,  
DEFINITION mRNA sequence.  
ACCESSION AA886109  
VERSION AA886109.1 GI:3001217  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 398)  
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900538.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,  
Rodrigo F. Chuangui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LINL at:  
[www-bio.linnl.gov/bdrp/image/image.html](http://www-bio.linnl.gov/bdrp/image/image.html)

Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 392.

FEATURES  
SOURCE

1. .398  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1274625"  
/clone\_1lb="NCI\_CGAP\_Pr12"  
/sex="male"  
/tissue\_type="metastatic prostate bone lesion"  
/lab\_host="DH10B"  
/note="Vector: PAMPI0; mRNA made from metastatic prostate  
lesion of the bone, cDNA made by oligo-dT priming.  
Non-directionally cloned. Size-selected on agarose gel,  
average insert size 600 bp. Library made by D. Krizman,  
NIH."

BASE COUNT 100 a 81 c 86 g 131 t  
ORIGIN

Query Match 21.6%; Score 335.8; DB 39; Length 398;  
Best Local Similarity 99.4%; Pred. No. 1,7e-76;  
Matches 337; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1212 ccagttaccgacatgcacacatctgattgctgacctgagatgagacagatgaaataat 1271  
|||||  
DB 42 CCACTTACCAGCATGACCAATCTGATTGCTGACCTGATGGAACAGAGTGAATAAT 101

QY 1272 gaattacaagaagattattacattcatctggtttagaattatattgcacaaagcaccac 1331  
|||||  
DB 102 GATTTACAAAGAAATTTACATTCATCTGTTAGACTTAATATGCGACACGCCACCAC 161

QY 1332 gaccttcccgagtgacacccgctcagctgcaatgagggcgctgctccatcaacgagggc 1391  
|||||  
DB 162 GACCTTCCAGGAGGACACCCGCTCAGCTGCAATGAGGGGCTGCTCATACCGGGGC 221

QY 1392 gctgtcccgacgacgagctgggctgagctgagatctgactctagctgagacagactcctg 1451  
|||||  
DB 222 GCTGTCCCGCAGCAGCTGCGGGCTGAGAGCTGAGACTGACTCTAGCAGAGTCTG 281

QY 1452 gtgtatgttttcagaagaatggttgaagttaagtgtttaactctgctcaattgtatgctcag 1511  
|||||  
DB 282 GTGTATGTTTTCAGAAATGCTTGAAGTATGTGTTAAATCTGCTCATTCGTATGCTAG 341

QY 1512 gtatacatatgatttcaataaataaacttttaaga 1550  
|||||  
DB 342 GTTATACATATGATTTCATTAATGAACCTTTTAAGA 380

RESULT 8  
AI140814/c 433 bp mRNA EST 13-FEB-1999  
LOCUS  
DEFINITION ool5c05.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
IMAGE:1566248 3, mRNA sequence.  
ACCESSION  
AI140814  
VERSION  
AI140814.1 GI:3648271  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
1 (bases 1 to 433)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

JOURNAL  
Unpublished (1997)

On Feb 10, 1998 this sequence version replaced gi:2339944.

COMMENT  
Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LINL; contact the  
IMAGE Consortium ([info@image.linnl.gov](mailto:info@image.linnl.gov)) for further information.

Insert Length: 958 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES  
SOURCE

1. .433  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1566248"  
/clone\_1lb="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker; Site 1: Not I; Site 2: Eco RI;  
Equal amounts of plasmid DNA from five normalized  
libraries were mixed, and ss circles were made in vitro.  
Following HAP purification, this DNA was used as tracer in  
a subtractive hybridization reaction. The driver was  
PCR-amplified cDNAs from pools of 5,000 clones made from  
the same 5 libraries. The pools consisted of the following  
libraries and clones: Soares NBHSF pool 1:  
303384-310919, 323208-325895 Soares NB2HP pool 1:  
145032-147335, 147720-148103, 148872-149255, 15002 -  
150407, 151176-152327 Soares NB2HP8-9W pool 1:  
758280-760583, 772104-774407 Soares NBHPA pool 1:  
304776-306311, 320136-322823, 326280-326653 Soares NBHOR  
pool 1: 723720-726407, 739080-740999 Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 119 a 91 c 88 g 135 t  
ORIGIN

Query Match 21.2%; Score 328.4; DB 42; Length 433;  
Best Local Similarity 99.4%; Pred. No. 1,4e-74;  
Matches 340; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1212 ccagttaccgacatgcacacatctgattgctgacctgagatgagacagatgaaataat 1271  
|||||  
DB 354 CCACTTACCAGCATGACCAATCTGATTGCTGACCTGATGGAACAGAGTGAATAAT 295

QY 1272 gaattacaagaagattattacattcatctggtttagaattatattgcacaaagcaccac 1331  
|||||  
DB 294 GATTTACAAAGAAATTTACATTCATCTGTTAGACTTAATATGCGACACGCCACCAC 235

QY 1332 gaccttcccgagtgacacccgctcagctgcaatgagggcgctgctccatcaacgagggc 1391  
|||||  
DB 234 GACCTTCCAGGAGGACACCCGCTCAGCTGCAATGAGGGGCTGCTCATACCGGGGC 175

QY 1392 gctgtcccgacgacgagctgggctgagctgagatctgactctagctgagacaga-ctcct 1450  
|||||  
DB 174 GCTGTCCCGCAGCAGCTGCGGGCTGAGAGCTGAGACTGACTCTAGCAGAGAGCTCCT 115

QY 1451 ggtatagttttcagaagaatggtcgtgaagttaagtgtttaactctgctcaattgtatgctc 1510  
|||||  
DB 114 GGTTATGTTTTCAGAAATGCTTGAAGTATGTGTTAAATCTGCTCATTCGTATGCTAG 55

QY 1511 ggtatacatatgatttcaataaataaacttttaaga 1552  
|||||  
DB 54 GGTATACATATGATTTCATTAATGAACCTTTTAAGAAA 13

RESULT 9  
AI038590/c 384 bp mRNA EST 28-AUG-1998  
LOCUS  
DEFINITION ox34e03.s1 Soares.Total\_Fetus\_NB2HP8\_9W Homo sapiens cDNA clone  
IMAGE:1658236 3', mRNA sequence.  
ACCESSION  
AI038590  
VERSION  
AI038590.1 GI:3277784  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;



DB	273	GGTGTATGTTTTCACAAATGCGTTCGAGTAAGTATGCTTTAAATCTGCTCATTTGCTATGCTA	332
OY	1511	ggtatcatatgatgttttccataataatgaacttttaaaga	1550
Db	333	GGTATACATATGATTTTCATTAATGAACCTTTTAAAGA	372
RESULT	12		
LOCUS	AA029201	439 bp mRNA	EST 09-MAY-1997
DEFINITION	zk12f08.s1 Soares_pregnant_uterus_NbHPV Homo sapiens CDNA clone		
ACCESSION	AA029201		
VERSION	AA029201.1	GI:1496605	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
AUTHORS	1 (bases 1 to 439) Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mawds, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.		
TITLE	Generation and analysis of 280,000 human expressed sequence tags		
JOURNAL	Genome Res. 6 (9), 807-828 (1996)		
MEDLINE	97044478		
COMMENT	On Sep 21, 1992 this sequence version replaced gi:279290. Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 2448 Std Error: 0.00 Seq primer: -40M13 fwd. from Amersham High quality sequence stop: 371. Location/Qualifiers 1..439 /organism="Homo sapiens" /db_xref="GDB:3756345" /db_xref="taxon:9606" /clone="IMAGE:470343" /clone_lib="Soares_pregnant_uterus_NbHPV" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: uterus; Vector: pT73-Pac; Site:1: Not I; Site:2: Eco RI; 1st strand CDNA was primed with a Not I - oligo(dT) primer [5', AACGGAAGAAGTTGCGCGCGCCCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."		
FEATURES			
SOURCE			
BASE COUNT	120 a 92 c 91 g 136 t		
ORIGIN			
Query Match	21.0%; Score 326.4; DB 27; Length 439;		
Best Local Similarity	99.4%; Pred. No. 4.8e-74;		
Matches 338; Conservative 0; Mismatches 1; Indels 1; Gaps 1;			
OY	1212	ccagttaccacgcatgcacacatctgattctgcagctgagtgagacagatgaataaat	1271
Db	25	CCAGTTACCCACGACATCCACATCTGATTTGCTGACGACTGATGAGAACAGAGTGAATTAAT	84

BASE COUNT	127 a	94 c	93 g	141 t	ORIGIN
Query Match	21.0%	Score 326.4;	DB: 28;	Length 455;	Went through one round of normalization. Library constructed by M. Patricia Bonaldo."
Best Local Similarity	99.4%	Pred. No. 4.8e-74;			
Matches 338;	Conservative	0;	Mismatches 1;	Indels 1;	Gaps 1;
QY 1212	ccagttaccacgcatctgcacacatcgtatctgtctgcacgtgatgcgaacagatgaataaat	1271			
Db 28	CCAGTTACCACGCAATGCCACAACTCGATTGCTGTGACCTGATGGAACACAGTGAATTAAT	87			
QY 1272	gaattacaagaagatatttaacatcattcgttattgagactaatatgcccacaagcaccac	1331			
Db 88	GATTTCACAAAGAGATATTTCATTCATCTGTGGTTTGAAGCTTAATATGCCAACAGCAGCAC	147			
QY 1332	gaccttcccaagggtgacacccgccttaagctctgcatgtggggctcgttctcataaagcgggc	1391			
Db 148	GACCTTCCACAGGGGACACCCGCTCAGCTCAGAGGGGGCTGGTCTCATCAGCCGGGC	207			
QY 1392	gcttccccgcgaacgaagtcgggctgtgagctgtgagctgtgacctaagcagaaga-ctctt	1450			
Db 208	GCTGTCCCCGACACGACGCGGGCTGGAGCTGAGTCTTACTCTTACTATGACAGAGACTTCT	267			
QY 1451	gggtatgttttcagagaatgtgcttgaagttaigtgtttaaactctgcatctcgatgcta	1510			
Db 268	GGGTGATGTTTTCAGAAATGGCTTGAAGTTATGTGTGTTAAATCTGCTCATTCGTATGCTA	327			
QY 1511	ggttatacatatgatattccaataaatgaacttttaaga	1550			
Db 328	GGTTATACATATGATTTTCAATTAATGAATCTTTTAAAGA	367			
RESULT 14					
LOCUS	AI168222	462 bp	mRNA	EST	23-OCT-1998
DEFINITION	OK30C01.x1 Soares_NSF_F8_9W_OT_PA_P_S1				Homo sapiens cDNA clone
ACCESSION	IMAGE:11509312	3',	mRNA sequence.		
VERSION	AI168222				
KEYWORDS	AT168222.1	GI:3701392			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
TITLE	Euthelia; Primates; Catarrhini; Hominiidae; Homo.				
JOURNAL	NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.				
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
	Unpublished (1997)				
	On Jan 19, 1998 this sequence version replaced gi:2284583.				
	Contact: Robert Strausberg, Ph.D.				
	Tel: (301) 496-1550				
	Email: Robert.Strausberg@nih.gov				
	This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.				
	Insert Length: 647 Std Error: 0.00				
	Seq primer: -40m13 fwd. ET from Amersham				
	High quality sequence stop: 386.				
FEATURES	location/Qualifiers				
SOURCE	1..462				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/clone="IMAGE:11509312"				
	/clone_id="Soares_NSF_F8_9W_OT_PA_P_S1"				
	/lab_host="DH10B"				
	/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from five normalized libraries were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in				



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 21:16:57 ; Search time 158.18 Seconds  
(without alignments)  
1245.804 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 1552  
Sequence: 1 cttctgagatcgcagagaa.....aaatgacttttaagaana 1552

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCNUS\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	90.6	5.8	7218	1 US-08-232-463-14	Sequence 14, Appl
2	42.4	2.7	543	4 US-08-728-323A-1	Patent No. 5273901
3	41.8	2.7	3489	4 US-08-770-379-20	Sequence 1, Appl
4	41.8	2.7	32207	3 US-08-257-073-4	Sequence 20, Appl
5	41.6	2.7	2223	2 US-08-257-073-4	Sequence 4, Appl
6	41.2	2.7	697	7 5171843-10	Patent No. 5171843
7	41.2	2.7	1137	7 5171843-8	Patent No. 5171843
8	40.4	2.6	234	3 US-08-469-802B-3	Sequence 3, Appl
9	40.4	2.6	234	3 US-08-267-803B-3	Sequence 3, Appl
10	39.6	2.6	9551	4 US-08-056-200-93	Sequence 93, Appl
11	39.6	2.6	9551	4 US-08-800-644-93	Sequence 93, Appl
12	38.8	2.5	195	2 US-08-469-802B-2	Sequence 2, Appl
13	38.8	2.5	195	2 US-08-267-803B-2	Sequence 2, Appl
14	38.8	2.5	2277	2 US-08-676-967-2	Sequence 2, Appl
15	38.8	2.5	2277	2 US-08-676-967-2	Sequence 2, Appl
16	38.8	2.5	2277	4 US-09-098-487-2	Sequence 2, Appl
17	38.8	2.5	4200	1 US-08-242-932-1	Sequence 1, Appl
18	38.8	2.5	4200	1 US-08-714-481-1	Sequence 1, Appl
19	38.8	2.5	4200	6 PCT-US95-06111-1	Sequence 1, Appl
20	38	2.4	325	3 US-08-332-766A-11	Sequence 11, Appl
21	38	2.4	1772	4 US-08-960-022-13	Sequence 13, Appl
22	37.4	2.4	171	2 US-08-469-802B-5	Sequence 5, Appl
23	37.4	2.4	171	2 US-08-267-803B-5	Sequence 5, Appl
24	37.2	2.4	3730	1 US-08-242-932-8	Sequence 8, Appl
25	37.2	2.4	3730	1 US-08-714-481-8	Sequence 8, Appl
26	37.2	2.4	3730	6 PCT-US95-06111-8	Sequence 8, Appl
27	37	2.4	1394	1 US-07-730-953-1	Sequence 1, Appl

## ALIGNMENTS

28	36.8	2.4	168	2 US-08-469-802B-4	Sequence 4, Appl
29	36.8	2.4	168	3 US-08-267-803B-4	Sequence 4, Appl
30	36.8	2.4	533	7 5482709-5	Patent No. 5482709
31	36.2	2.3	43795	5 US-08-742-185-101	Sequence 101, App
32	35.8	2.3	1709	4 US-09-010-398-2	Sequence 2, Appl
33	35.8	2.3	4363	3 US-08-685-576-5	Sequence 5, Appl
34	35.8	2.3	7210	3 US-08-257-963B-10	Sequence 10, Appl
35	35.8	2.3	7210	6 PCT-US95-07201-10	Sequence 10, Appl
36	35.8	2.3	22481	6 PCT-US95-07201-43	Sequence 43, Appl
37	35.6	2.3	2729	1 US-08-412-431-2	Sequence 2, Appl
38	35.6	2.3	2729	1 US-08-623-679-2	Sequence 2, Appl
39	35.6	2.3	2729	5 US-08-933-774-2	Sequence 2, Appl
40	35.4	2.3	1647	7 5405943-3	Patent No. 5405943
41	35.2	2.3	597	3 US-08-332-766A-19	Sequence 19, Appl
42	34.8	2.2	1535	5 US-08-747-574-3	Sequence 3, Appl
43	34.8	2.2	3376	1 US-08-320-559-29	Sequence 29, Appl
44	34.8	2.2	3376	6 PCT-US94-04496-29	Sequence 29, Appl
45	34.6	2.2	154	2 US-08-469-802B-6	Sequence 6, Appl

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZpt-Fls  
US-08-232-463-14





Db 2054 AGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGAGAGCAGCAGATG 2113  
QY 605 tggaaaagttaccagaagaaagattcagctcttaactgaggaattcaagcc 664  
Db 2114 AGCAGAGAGCAGAGAGGATGAGCAGAGCAGCAGCAGATGAGCAGAGATGAGCAGCAGC 2173  
QY 665 agctggcttcctggggaatgagtgatcacaaggtgctgaggaatgagctatcagctga 724  
Db 2174 AGCAGAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAG 2233  
QY 725 ataaaccacacatgagaaagatgagcagaagagcagagagctcagaagcaaaa 784  
Db 2234 ATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2293  
QY 785 ctaacaggatcttgaatctaaagatcaggaatagagaattgagaatagaactgagatg 844  
Db 2294 AGCAGCAGCAGATGAGCAGCAGCAGCAGCAGATGAGCAGATGAGCAGCAGCAGATGAGCAG 2353  
QY 845 aaagcaaaacacactgagacagagcagagagcagccctggccagagagagtgcc 904  
Db 2354 AGCAGAGCAGCAGGATTAGAGAGCAGCAGCAGCAGATGAGAGAGCAGCAGAGATTAG 2413  
QY 905 tgaactaacagaaactgctgggagaaactcagcaccacactgcaacctcaccagatcga 964  
Db 2414 AGGAGCAGCAGCAGAGGATTAGAGAGCAGCAGCAGCAGCAGATGAGAGAGCAGCAGAGAT 2473  
QY 965 tagctcaactcagtcagaagaaaggtatataatgaggaatgaggaatgagga 1024  
Db 2474 TAGAGAGCAGCAGCAGCAGGATTAGAGAGCAGCAGCAGCAGATGAGAGAGCAGCAGAGTGG 2533  
QY 1025 gaatagaagaattgagagagcagtgctcagcagatgagagagagagagagagagagag 1084  
Db 2534 AAGAGCAGCAGCAGCAGGAGGAGGAGAGCAGCAGCAGCAGCAGAGAGCAGGAAATTAGAGG 2593  
QY 1085 aaaggtcgaagcagctgagtaagcagcagcagcagcagcagcagcagcagcagcagcagc 1144  
Db 2594 AGGTGAGAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 2653  
QY 1145 tcagcaagcagaagcagcttctcctgagagagcagagcctgctgcgaagagagtg 1197  
Db 2654 AAGAGCAGCAGCAGCAGCAGGATTAGAGAGGAGTGAAGCAGCAGAGCAGAGAG 2706

RESULT 4  
US-08-770-379-20/c  
; Sequence 20, Application US/08770379  
; Patent No. 5849564  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Yuan  
; APPLICANT: Bohenzky, Roy A.  
; APPLICANT: Russo, James J.  
; APPLICANT: Edelman, Isidore S.  
; APPLICANT: Moore, Patrick S.  
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Cooper & Dunham LLP  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/770.379  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 52342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 2.7%; Score 41.8; DB 3; Length 32207;  
Best Local Similarity 40.9%; Pred. No. 0.18;  
Matches 316; Conservative 0; Mismatches 457; Indels 0; Gaps 0;  
QY 425 accagttgagaaagagctgagagagcagcagcagcagcagcagcagcagcagcagcagcagc 484  
Db 20063 AGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 20004  
QY 485 agcaagaagaagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 544  
Db 20003 AGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19944  
QY 545 agtacttgatcaaatgattgattcttctcagataatgcccacagagagccag 604  
Db 19943 AGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19884  
QY 605 tggaaaagttacaagaagaagattcagctataactcagaggaatcaagcc 664  
Db 19883 AGCAGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19824  
QY 665 agctggcttcctggggaatgagtgatcacaaggtgctgaggaatgagcagctatcagctga 724  
Db 19823 AGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19764  
QY 725 ataaaccacacatgagaaagatgagcagaagagcagcagagagagagagagagagagagagag 784  
Db 19763 ATGAGCAGCAGCAGCAGCAGGATTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 19704  
QY 785 ctaacagagatcttgaatataagatcaggaatagagaatgagaatgagaactgagatg 844  
Db 19703 AGGAGCAGGATTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 19644  
QY 845 aaagcaaaacacactgagaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 904  
Db 19643 AGCAGAGCAGCAGGATTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 19584  
QY 905 tgaactaacagaactgctgggagaaactgagcagcagcagcagcagcagcagcagcagcagcagcag 964  
Db 19583 AGGAGCAGCAGCAGGATTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 19524  
QY 965 tagctcaactcagtcagaagaaaggtatataatgaggaatgaggaatgaggaatgaggaatgagga 1024  
Db 19523 TAGAGAGCAGCAGCAGGATTAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 19464  
QY 1025 gaatagaagaattgagagagcagtgctcagcagatgagagagagagagagagagagagagag 1084  
Db 19463 AAGAGCAGCAGCAGGAGGAGGAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 19404  
QY 1085 aaaggtcgaagcagctgagtaagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1144  
Db 19403 AGGTGAGAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTGG 19344  
QY 1145 tcagcaagcagaagcagcttctcctgagagagcagagcctgctgcgaagagagtg 1197  
Db 19343 AAGAGCAGCAGCAGCAGGATTAGAGAGGAGTGAAGAGCAGCAGAGCAGAGAG 19291

```
RESULT 5
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 576597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-4

Query Match      2.7%; Score 41.6; DB 2; Length 2223;
Best Local Similarity 56.6%; Pred. No. 0.045;
Matches 77; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO: 10:
; LENGTH: 697
; 517843-10

Query Match      2.7%; Score 41.2; DB 7; Length 697;
Best Local Similarity 44.1%; Pred. No. 0.031;
Matches 267; Conservative 0; Mismatches 333; Indels 6; Gaps 2;
```

```
QY 601 caggtgaaaggtacaaagaaagattcaagctataaactgaggaattcaa 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 cagaaccaaataatccacgtgaataagctgaagcaacgagacagcagatgac 61
QY 661 agccagctgcttcggaatgtgatccaaagtggtgagaatcgtctatcag 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 agccagcagagacagagcagatgacagcagcaggtgatalagacatgacaacag 121
QY 721 ctgaataaaccaacatggaga---agatgagcagaagaaggcagagattcaga 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 caggaacataagcgcgtggaacacagcagagatagagcagatggaacgacagag 181
QY 778 gcaaaactaacagggatcttgaaattaaagatcaggaatagagaaattgagaatgaa 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 acagagcagatggaacagcagcagagagacagagcagatggaacacagcagagagag 241
QY 838 ctgtgaaagcaaacacacacttggaaacagagcagcagagagcagccctgccaagag 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 cagatggaacacacagcaggtgatagagcgtggacaacacagcaggtgatagcagctg 301
QY 898 gagtgcctgagactaacagacactgtggcgatctgtgacccaactgcacctcacaga 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 302 gacaacccagcagacatagagcagatggaacccagcagcagatagagcagctggacagc 361
QY 958 tctgaatagctcaactcagtcagaaagaaagatataatgataattggaaagtta 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 cagcagagatagagcagatggaacgacagcagagatagagcagctggacagcagcag 421
QY 1018 caggaagaagaatgaaatctggaagaacagtgtgtccagcatgggagatcagagagc 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 422 gagatagagcagatggaacagcagcagcagatagagcagctggacagcagcagaga---g 478
QY 1078 atgagcaaaagcctaagcagctgataagcacagccagcagcagcagcagcagctgtg 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 479 atagagcagcctggacagcagcagcagagatagacagcctggacagcagcagagatagag 538
QY 1138 cagctcctcagcaagcagaaacagcttctcctggaagagcagagcctgtggaagagtg 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 539 cagctggaacacagcagcagaaatgtgcaagtggaacagcagcagcagagaaacgacagag 598
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1198 gaccgg 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 599 gacag 604
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
517843-8
; Patent No. 5171843
; APPLICANT: NUSSENZWEIG, VICTOR
; TITLE OF INVENTION: IMMUNOGENIC POLYPEPTIDE AND METHOD FOR
; PURIFYING IT
```

```
;
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/175,112
; FILING DATE: 30-MAR-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 754,645
; FILING DATE: 9-JUL-1985
; APPLICATION NUMBER: 115,634
; FILING DATE: 26-OCT-1987
; APPLICATION NUMBER: 649,903
; FILING DATE: 12-SEP-1984
; SEQ ID NO: 8
; LENGTH: 1137
5171843-8

Query Match      2.7%; Score 41.2; DB 7; Length 1137;
Best Local Similarity 44.1%; Pred. No. 0.04;
Matches 267; Conservative 0; Mismatches 333; Indels 6; Gaps 2;

QY 601 caggtggaagaaagttacaaagaagaatttcagctattatcaacttgagaagaattcaa 660
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 cagaacccaataatccacgtgaaataaagctgaagcaaccagagagacagagcagatgagac 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 661 agccagctgcttcggaagaatgagatgcacaagaagtgtggaagaatgcgtatcacg 720
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 308 agccagcagagaaacagagagatgagacagccagcaggtgatagagacagatgagacaaccag 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 721 ctgaataaaacaaactgagaga--aggaatgagcagaagaagagagagagagagagagag 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 368 cagagatagagcagctgagacacacagcagagatagagcagatgagacagcagcagcagag 427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 778 gcaaaaactaaacagagagatcttaaatgaatcaggaatagagaatgagagaatagaga 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 428 acagagcagatgagaaagcagcagagagagacagagcagatgagacacacagcagagagaga 487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 838 ctgagtgaaagaacaaacactcttgagacagagagcagagagcagccctgscagagagag 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 488 cagatgagacaaccacagctgagtaagcagctgagacaaccacagctgagtaagcagctg 547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 888 gagtgcttgagactaacaagaactgtggtggaatctgagacacaaactgacactcacaga 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 548 gacaacccagcagagatagagcagatgagcagcagcagcagagatagagagagctgagcagc 607
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 958 tctggaataagctcaactcagctcaagaaaaaaaggtatataatgagaaatgaggaagtta 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 608 cagcagagagatgagacagatgagacagcagcagcagagatagacagctgagacagcagcag 667
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1018 cagagaaagaatgagaagatgtgagagaaacagctgtgtccagcagtgagagagtagatagagacg 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 668 gagatagagcagatgagacagcagcagagagatagacagctgagacagcagcagaga--g 724
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1078 atgaagcaaaagctlaagcagcttgataagcaacagcagcagcagcagcagcagctgtgtg 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 725 atagagcagctgagacagcagcagcagagatagagcagctgagacagcagcagagagatagag 784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1138 cagctcctcagcagcagagaacacagcttctctcgtgagagagcagagcctgtgcaagaagtg 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 785 cagctgagacagcagcagagaatgtgtcaggtgagacagcagcagcagagagaagaacagcagag 844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1198 gacccg 1203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 845 gacagag 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
US-08-469-802B-3
; Sequence 3, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P. W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Chung, Ming-Yi
```

```
;
; APPLICANT: zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muelling, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muelling, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-469-802B-3

Query Match      2.6%; Score 40.4; DB 2; Length 234;
Best Local Similarity 50.5%; Pred. No. 0.028;
Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 1038 ggaagaaagctgtgtccagcagtgagagagtagatagagcagatgagaagaagctaagca 1097
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 30 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1098 gctgataagcacaagcagcagcagcagcagcagcagctgtgtcagctctcagcaagcagaa 1157
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 90 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1158 ccagcttctctgtggaagcagagcctgtcgaagaagtggaacggctgtcgagccagtt 1217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 150 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1218 acccagcagtcac 1231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 210 CCTCAGCAGGCTC 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
US-08-267-803B-3
; Sequence 3, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P. W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
```



```

; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,644
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/056,200
; FILING DATE: 30-APR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.
; REGISTRATION NUMBER: 36,799
; REFERENCE/DOCKET NUMBER: NIH054.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (714) 760-0404
; TELEFAX: (714) 760-9502
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9551 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1507..1644
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 1645..2511
; NAME/KEY: CDS
; LOCATION: 2512..8070
; US-08-800-644-93

Query Match      2.6%; Score 39.6; DB 4; Length 9551;
Best Local Similarity 51.1%; Pred. No. 0.37;
Matches 93; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 1028 atgaagaattgaggaacagtggtccacgacatggagagtgactgagacgatagaacaaa 1087
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3089 acgaggaataaagaaagagtgagagacgcgacgacagctccggaagaaagagaga 3148
QY 1088 ggcctaagagcagctggaataagcagccagccagccagcagctggtgagcagctctca 1147
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3149 acttgacagaaagagagcgccgacgcgcaaaagagagctccagagagaaagagacacttac 3208
QY 1148 gcaagcagaacacagcgttctctcgtgagagcagaagcctgtcgcgaagaagagtgagccgctgc 1207
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 3209 ggaagctgagcgagcgaagagctgagagagagcagcgccgacgagaaagagacagcaccacaa 3268
QY 1208 gg 1209
    ||
DB 3269 gg 3270

RESULT 12
US-08-469-802B-2
; Sequence 2, Application US/08469802B
; Patent No. 5741645
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P.W.
```

```

; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5741645
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muelting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: 119 No. 5741645th Fourth Street, Suite 203
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,802B
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muelting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110.00030101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1225
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 195 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-469-802B-2

Query Match      2.5%; Score 38.8; DB 2; Length 195;
Best Local Similarity 51.1%; Pred. No. 0.07;
Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 1054 cagcagtgagagatcatgagcagatgataagcaagcagcagctgagataagcagc 1113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 66
QY 1114 cagccacagccagccagcagctggtgagctctcagcaagcagcagcagcttctcggag 1173
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 126
QY 1174 agcagagcctgtcgaagaagtgagtcgagcagcagcagcagcagcagcagcagcagcagc 1231
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 127 CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 184

RESULT 13
US-08-267-803B-2
; Sequence 2, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
; APPLICANT: Orr, Harry T.
; APPLICANT: Rannum, Laura P.W.
; APPLICANT: Chung, Ming-Yi
; APPLICANT: Zoghbi, Huda Y.
; TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
; Patent No. 5834183
; TITLE OF INVENTION: Type 1 and Method for Diagnosis
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Muelting, Raasch, Gebhardt & Schwappach, P.A.
; STREET: P.O. Box 581415
; CITY: Minneapolis
```

STATE: MN  
COUNTRY: USA  
ZIP: 55458-1415  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/267, 803B  
FILING DATE: 28-JUN-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: McCormack, Myra H.  
REGISTRATION NUMBER: 36,602  
REFERENCE/DOCKET NUMBER: 110.00030120  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-305-1217  
TELEFAX: 612-305-1228  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 195 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
US-08-267-803B-2

Query Match 2.5%; Score 38.8; DB 3; Length 195;  
Best Local Similarity 51.1%; Pred. No. 0.07;  
Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Oy 1054 cagcgtggagagatcctcagagacgttgaaagcaagcctaagcgacgtgataagacagc 1113  
Db 7 CAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 66  
Oy 1114 cagcgcaagccagcagcagcgtgtgagctctcagcaagcaagcagcttctctggag 1173  
Db 67 CAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 126  
Oy 1174 agcgagagcctctcgaaagtgtagccggtctcgagaccagttaccagcatgtcac 1231  
Db 127 CAGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 184

RESULT 14  
US-08-676-967-2

; Sequence 2, Application US/08676967  
; Patent No. 5747317  
; GENERAL INFORMATION:  
; APPLICANT: COLLINS, KATHLEEN  
; TITLE OF INVENTION: Human Telomerase  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Science & Technology Law Group  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676, 967  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard A  
; REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-676-967-2

Query Match 2.5%; Score 38.8; DB 2; Length 2277;  
Best Local Similarity 25.6%; Pred. No. 0.28;  
Matches 99; Conservative 72; Mismatches 216; Indels 0; Gaps 0;

Oy 466 gaaagaagacttgatctcagcaagagaaagggccattgagaagacatgatgaaaag 525  
Db 619 GAAARWMSNCAYGARWMSNAARCAVCARGARWMSNGTNAARAAARAGWNGARGAR 678  
Oy 526 gaataacgaaagaaagggatgacatgagatcgaagatgtgattcttcagaatatt 585  
Db 679 GATATGARGARAGARAAAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGAYGNGTN 738  
Oy 586 gcccaactggagggcccggttgaaagggttaaaaggaagaaagatttcagctataatcaa 645  
Db 739 TTYGAYGAYGARAYGARAYGARARARAYTHGARWMSNAARGTNACNAARCCNGTNGAR 798  
Oy 646 ctggaggaattcaaaagcagctgtcttcgggaatgagatgtacaaggtgtgtga 705  
Db 799 ATHCARARWMSNGCNGTNAARWMSNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNGCNG 858  
Oy 706 gaatgtcgtatcagctcgtgaataaaccacatgagaaagatggagcgaagaaagagcac 765  
Db 859 GAYWMSNGAYTYNGARWMSNGAYWMSNTHGAYGAYGNGGARAYTYNGCNCARWMSNGAY 918  
Oy 766 agagagttcagagcaaaactaacagagatcttgaatlaagatcaggaatagagaa 825  
Db 919 ACNMSNACNGARGAR 978  
Oy 826 ttgagaatagacttgatgaaagcaaa 852  
Db 979 YTNCCMSNGAYGTMAAYGARGNAR 1005

RESULT 15  
US-08-676-974-2

; Sequence 2, Application US/08676974  
; Patent No. 5770422  
; GENERAL INFORMATION:  
; APPLICANT: COLLINS, KATHLEEN  
; TITLE OF INVENTION: Human Telomerase  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Science & Technology Law Group  
; STREET: 268 Bush Street, Suite 3200  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/676, 974  
; FILING DATE:  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Osman Ph.D., Richard A

REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-676-974-2

Query Match 2.5%; Score 38.8; DB 2; Length 2277;  
Best Local Similarity 25.6%; Pred. No. 0.28;  
Matches 99; Conservative 72; Mismatches 216; Indels 0; Gaps 0;

OY 466 gaaaagaactgcacatcagcaagaagaaaggccattgagaagacatgatgaaag 525  
DB 619 GAAARWSNCATGARWSNARCAICARGARWSGTVNAPARAARAGNNGARGARGAR 678  
OY 526 gaataacgaagaaggaggtacatggatgataagattgtctcagaatatt 585  
DB 679 GAYATGARARGARGARAAIYGAICYGAYGAYGAYGAYGARGARGARGARGAYGNGTN 738  
OY 586 gcccaactggagccaggtggaagaagttacaaggaagaagattcagctaatca 645  
DB 739 TTYGAYGAYGARGARGARGARGARAAIYTHGARWSNARGTNANARCCNGTNCAR 798  
OY 646 ctggagggaattcaagccagctgtctctcgggaatgcatgcacaaggtgtgtga 705  
DB 799 ATHCARARARMGNCGTNARMGNCNCNCNCNCAARMSWSNGAYCAYWSNGARGAR 858  
OY 706 gaatgcctatcagctgataaaccacacatggagaagatgagcagaagaagagcac 765  
DB 859 GAYWSNGAYITNGARGARMSNGAYMSNATHGAYGAYGCGARGARTTNGCNCARMSNGAY 918  
OY 766 agagagttcagagcaaaaactacagaggtcttgaattaaagatcaggaataagagaa 825  
DB 919 ACNWSNACNGARGARGARGARGAYARCGCNGTNCARGTWSNAAYARAARAARBMGNAAR 978  
OY 826 ttgagaatagaactggtatgaagcaaa 852  
DB 979 YTNCCNWSNGAYGTNAAYGARGGNAAR 1005

Search completed: May 20, 2000, 00:27:48  
Job time: 11451 sec

**THIS PAGE BLANK (USPTO)**



1	1305	84.1	1789	1	X40059	Colon cancer assoc
2	1277.4	82.3	2353	1	X400542	Homo sapiens secret
3	852	54.9	852	1	X40054	Colon cancer assoc
4	279.4	18.0	310	1	V89187	EST clone CH699. N
5	114.4	7.4	317	1	T24615	Human gene signatur
6	113.8	7.3	432	1	X51883	Human secreted pte
7	67.2	4.3	1666	1	Q87587	DNA encoding leucoc
8	57.2	3.7	4000	1	T91902	Mannose-1-phosphat
9	52.2	3.4	1435	1	X32489	Seq ID No:4 used i
10	52.2	3.4	2100	1	V42601	Streptococcus uber
11	46	3.0	1908	1	N71064	Gene encoding Plas
12	44.8	2.9	6755	1	V21511	Staphylococcal bac
13	44.2	2.8	3359	1	T05668	Chicken leucocytoz
14	44.2	2.8	3566	1	T46187	DNA encoding cell
15	44.2	2.8	3624	1	T46186	DNA encoding cell
16	43.8	2.8	2640	1	V04529	Human OC-116 kDa c
17	43.2	2.8	1048	1	X32486	Seq ID No:1 used i
18	43	2.8	110000	1	V21209_05	Continuation (6 of
19	42.4	2.7	563	1	Q23092	Antigen tc-7a gene
20	42	2.7	3353	1	X24681	Human synaptomemat
21	41.8	2.7	202	1	T23845	Human gene signatur
22	41.8	2.7	32207	1	V73805	KSHV LTR DNA (nucl
23	41.8	2.7	137507	1	V15941	KSHV long unique c
24	41.6	2.7	932	1	V53540	DNA encoding a PliA
25	41.6	2.7	2223	1	Q80808	Plasmodium falcipar
26	41.2	2.7	1137	1	Q33061	Plasmodium vivax c
27	41	2.6	724	1	X39663	Renal cancer assoc
28	41	2.6	862	1	X39858	Gastric cancer asso
29	40.6	2.6	2010	1	N91379	Intron 1 from huma
30	40.4	2.6	700	1	X88332	Spinoecerebellar at
31	40.2	2.6	734	1	V88907	EST clone IA200. N
32	40	2.6	2223	1	Q29187	CDNA encoding Plas
33	39.2	2.5	1258	1	X40056	Colon cancer asso
34	39.2	2.5	2277	1	V13634	Homo sapiens ambig

35	39.2	2.5	3653	1	X20280	<i>Borrelia burgdorferi</i>
36	39	2.5	6644	1	X33181	Base sequence of t
37	39	2.5	7372	1	X33182	Base sequence of t
38	39	2.5	7797	1	X33180	Complex virus bsr f
39	39	2.5	7996	1	X33184	Base sequence of t
40	38.8	2.5	195	1	084831	Spino cerebellar at
41	38.8	2.5	1397	1	T84935	Human prostate pro
42	38.8	2.5	1797	1	T84931	Human prostate pro
43	38.8	2.5	2277	1	V05370	Human telomerase p
44	38.8	2.5	3492	1	V13348	DNA encoding C-pet
45	38.8	2.5	4200	1	T03190	Group B Streptococ

## ALIGNMENTS

RESULT	1
AC	X40059
ID	X40059 standard; DNA; 1789 BP.
DT	02-JUL-1999 (first entry)
NC	Colon cancer associated gene.
KW	Cancer associated antigen; diagnosis; research; treatment; human.
KW	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
OS	prostate cancer; ss.
OS	homo sapiens.
PN	W09J04263-A2.
PD	28-JAN-1999.
PF	15-JUL-1998; U14679.
PR	22-JUN-1998; US-102322.
PR	17-JUL-1997; US-896164.
PR	10-OCT-1997; US-061599.
PR	10-OCT-1997; US-061765.
PR	10-OCT-1997; US-948705.
PR	11-OCT-1997; GB-021697.
PA	(LUDW-) LUDWIG INST CANCER RES.
PI	Chen Y, Gout I, Gure A, O'Hare Y, Old LJ,
PI	Pfeundtschuh M, Sahin U, Scanlan MJ, Stockert E,
PI	Thureci O.
DR	WPI: 99-332448/11.
PT	New isolated cancer associated nucleic acids and polypeptides -
PT	isolated using sera from cancer patients, used to develop products
PT	for the diagnosis, monitoring or treatment of cancers
PS	Claim 67: Page 660: 787pp: English.
CC	The invention relates to a method for diagnosing a disorder characterised
CC	by expression of a human cancer associated antigen precursor coded for by
CC	a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC	biological sample isolated from a subject with an agent that specifically
CC	binds to the NAM, an expression product or a fragment of an expression
CC	product complexed with an HLA molecule; and (b) determining the
CC	interaction between the agent and the NAM or the expression product as a
CC	determination of the disorder. The products and methods can be used in
CC	the diagnosis, monitoring, research, or treatment of conditions
CC	characterised by the expression of various cancer associated antigens.
CC	The invention provides nucleic acid sequences and encoded polypeptides
CC	which are cancer associated antigen precursors expressed in human breast
CC	cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC	lung cancer.
CC	Sequence 1: 789 BP; 622 A; 341 C; 452 G; 374 T;
SO	Sequence 1: 789 BP; 622 A; 341 C; 452 G; 374 T;
QY	1 cctctgtagtcacccgagaagcctaaacttacttattgaggaagaagtgtgaattgagaa 60
DB	1 ctttcgtgattgcattccgagacctaataaacttattgaggaagaagaagtgtgaattgagaa 60
OY	61 tcccatgaagattttttagagacgaactagttgatatcagaagaactgtgaagctt 120
DB	61 tcccatgaagattttttagagacgaactagttgatatcagaagaactgtgaagctt 120



CC activities.	2353 BP;	826 A;	451 C;	580 G;	496 T;
Sequence					

Query match	82.38;	Score 1277.4;	DB 1;	Length 2353;
-------------	--------	---------------	-------	--------------

Best Local Similarity 86.5%; Pred. No. 0;  
Matches 1536; Conservative 0; Mismatches 1; Indels 238; Gaps 2;

OY	16	gagaaagctaaaaacttacttrtcttgagaaaagatgtgaaattgagaaatcccaattgaagtt	75
Db	567	GAGAAAGCTAAAACTTACTTATGAGAAAAGTGGAAATTTGAGAAATCCCAATTGAAGTT	626
OY	76	ttgagaaagcactttagcttgataatacagagaacttgtgaaagactcttaaaagacaataaag	135
Db	627	TTGAGGAACGACTTAGCTGAATATTCAGAGAACTGTGGAGATCTTAAAGAGCACTAAAG	686
OY	136	cataaagaattctctctgctctctaactgtgaaacgltgtgtgtgtgtcttltttgaa	195
Db	687	CATTAAGAAATTTCTTCGGCTGCTTAATACTGTGAACCGGTGGGGCTTTGTTGGAA	746
OY	196	tgtgtcaagataaagctgtctctctcccaaacccataactaatggtcatatgacaagcctc	255
Db	747	TGTGCTCAGCATGAAGCTGTTCTTTCCCAACCCATACATTAATGTTCAATATGCAAGCAATC	806
OY	256	gaaagacttggtltaaaagaagagaatgacttgatgtctgcactggttccgtlaagagcagc	315
Db	807	GAAGACTGGTTAAAGAAAAGATGACTTGATGTGTGCACGTAGTTTCCGTAAAGGAGCAGC	866
OY	316	ttggcagatcgcagcaagaagaagaagtgtctatgaaacaggtgaaacaagltttgcaa	375
Db	867	TTGGCAATACGCGAGCAAAAGAGAGCAAGCTGTTATGAACAAGTGAACAAAGTTTGGCA	926
OY	376	atactctgaagaagccaattttgaaaaaaccaagccttaactcaagttgtgaccaagttagg	435
Db	927	ATATCTCAGGAGAGCCATTTTGAHAHAACCAAGCTTTATCTCAATGTGACCAAGTTGAGG	986
OY	436	aagagacttgtagagagcagagcgcgagcagcttgaaaaaaacttgcatctcagcaagaagaa	495
Db	987	AAGGAGCTGGAGAGCGAGCGGGGAGCGACATTGAHAHAAGAACTTGATCTGCAGCAAGAGAA	1046
OY	496	agggcacttgagaaagacatgatatgaaaaaagaaataacgaagaagaaggtatcatgga	555
Db	1047	AGGGCACTTTAGAAAAGCAATGATGAHAHAAGAAATTAACAAHAHAAGAAAGGAGTACATGTGGA	1106
OY	556	tcaagaatgttgaatctgtgtctcagaatattgcccaactgtaggcccgaagtgtgaaaagtt	615
Db	1107	TCAAGATGTTGATCTTGTCTCAAAATATTGGCCCACTGAGAGCCCAAGTGAHAHAAGTT	1166
OY	616	acaaagaagaaaagatttcaagctataatacaacttgagaaatccaagcagctgtgctct	675
Db	1167	ACAAAGGAAAAGATTCAGCTATTATTAACAATGGAGGAATTCAAAGCAGCTGGCTCT	1226
OY	676	cgggaaatgtagtccacaagaagtgtgtgtgagaatavgcgtatcagctgtaataaaacaac	735
Db	1227	CGGAAATGTGATGTCACAAAAGGTGTGTGGAGAAATCCGTATCACCTGAATTAHAACCAAC	1286
OY	736	atggagaagaagttagagcaagaagaagagcacaagagtttaagagcaaaaaactcaacagagat	795
Db	1287	ATGGAGAGAGATGAGCAHAHAHAAGGAGCACAGAGATTTGAGCAHAHAACCTAATACAGGGAT	1346
OY	796	cttgaatatlaaagatcagagaatcagaagaatttgagatatgaacttgatgtaagaagcaaa	855
Db	1347	CTTGAATTTAAAGATCAGGAATATAGAGAAATTGAGAAATTAAGAACTGGATGAHAHAACAA	1406
OY	856	caacttggaaacagagacagcagaagcagccctggccagagagaggtgtccctagactaa	915
Db	1407	CACTTGGAAAGAGGACAGCGAAGAGCGACGCCCTGGCCAGAGAGAGATGGCTGAGACTAACA	1466
OY	916	gaactgttgagcgaactctgagcaacaactggaacctc-----	952
Db	1467	GAACTGCTGGCGCAATCTTGAGCACCAACTGTACACTTACCAGACAGAAHAHAAGATTAACATT	1526
OY	952	-----	952

Db	1527	CAGCAGACGTTTAGCAAGCAGCAAGCAAAAGCCCAAGCCCTTCAGGCCACGCAAAAGACGACG	1586
Qy	952	-----	952
Db	1587	GAGCTGACACAGAAAGTACAGCAAAATGSAAGCCGACACTGACAAAAGTGAAAATGAACAG	1646
Qy	952	-----	952
Db	1647	TATTTGTGCTGACCTCCAGAAATACATTTTGTGACAAAGTTAAAGAGAAATGCTGTACA	1706
Qy	952	-----	952
Db	1707	TTAGCCAAAGAACTGSAACAAATCTCTCAAAAAGCCAGATCTGTAATTAAGCTACACTCACT	1766
Qy	979	caagaaaaaaggatatacatatgataaattcgggaagttacagagaagaatgaaagtac	1038
Db	1767	CAAGAAAAAAGGTATCATATGATGTAATTTGGGAAAGTTACAGAGAGAAATGAAGAATTG	1826
Qy	1039	ggggaaacaatgctgctcgaagatcgggaagatgaaatgaaagatgaaagcgtcaagcg	1098
Db	1827	GAGGAACACTGTGTCCAGCATGGGAGATGACATGAAACATGAAAGCAAAAGCTTAAAGGCA	1886
Qy	1099	ctgataaagacacagccaagccaagcccaagccaagctgctgcaagctctcgaagcgaag	1158
Db	1887	CTGGATTAAGACACAGCCAGGCCACAGCCGACAGCTGTGCTGACGCTCTCAGCAAGCAGAAC	1946
Qy	1159	cagctctcctcggagagagcagagcctgtcggaaaggatgaaaccgctcggaccagttta	1218
Db	1947	CAGCTTCTCTGTGGAGAGGACGAGACCTGTGGAAGAGGTGGACCGCGCTCGGACCCAGTTA	2006
Qy	1219	ccagagatccacaatctgattgctgactgtgataggaaacagttgaataataatgaattac	1278
Db	2007	CCAGGATCCCAAACTGATGCTGACTGCTGATGGAGAAAGATGAATTAATGATTATAC	2066
Qy	1279	aaagataattacattcatctctgtttagacttaatatgcacaagcaccacgacttc	1338
Db	2067	AAAGATATTTAATCAATTCATCTGTGGTTTAGACTTAAATATGCGCAACAGCACAGACTTC	2126
Qy	1339	ccaaggtgacaccgcctcagcctcagcggatggggctgtgctcctcatcaacgcgggcgtgtcc	1398
Db	2127	CCAGGTGTGACACCGCCTCAGCCTGCGAGTGGGGCTGTGCTCTCATCAACGCGGGCGTGTCC	2186
Qy	1399	ccgcaagcagctggggtctgagcctggagctgtgactcagctgagcaga-ctccgggtgatt	1457
Db	2187	CCGACAGCATGCGCTGGAGCTGGAGTGTGATCTTACGTGACCAACAGCTCTCGGTGTAT	2246
Qy	1458	gtttcagaatcggcttgaagtatcgtttaactgtctcatctcgtcatgtagttata	1517
Db	2247	GTTTTCAGAAATGGCTTGAAGTATGTGTTTAAATGCTGCATATGCTATGCTAAGTTATA	2306
Qy	1518	catatgatttcaataaataatgaactcttttaagaaa	1552
Db	2307	CATATGATTTTCATATTAATGAACCTTTTAAAGAAA	2341
RESULT	3		
X40054			
ID	X40054	standard; DNA; 852 BP.	
AC	X40054;		
DC	02-JUL-1999	(first entry)	
DE		Colon cancer associated gene.	
KW		Cancer associated antigen; diagnosis; research; treatment; human;	
KW		breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
OS		prostate cancer; ss.	
KS		Homo sapiens.	
PN	WO9004265-A2.		
PD	28-JAN-1999.		
PF	15-JUL-1998:	U14679.	
PR	22-JUN-1998:	US-102322.	
PR	17-JUL-1997:	US-896184.	
PR	10-OCT-1997:	US-061599.	
PR	10-OCT-1997:	US-061765.	

PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; GB-021697.  
 PA (LUDM-) LUDMIG INST CANCER RES.  
 PI Chen Y, Gout I, Sure A, Ohare M, Obata Y, Old LJ,  
 PI Pfeundschuh M, Sahlin U, Scanlan MJ, Stockert E,  
 PI Tureci O;  
 DR WPI; 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Claim 67, Page 650, 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 SQ Sequence 852 BP; 281 A; 178 C; 225 G; 168 T;

Query Match 54.9%; Score 852; DB 1; Length 852;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-219;  
 Matches 852; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 701 gtgtgagaatcgctatcagctgaataaacaacatgagaagaatgagacagaagaag 760  
 DB 1 GTGGAGAAATGCGCTTACAGCTGTAATTAACCAACATGAGAGAGATGAGCAAAAAGG 60  
 QY 761 agccacagaagctcagagcaaaaactaacagagatcttgaataataagatcagaagaatag 820  
 DB 61 AGCACAGAGAGATTCAGAGCAAAAACCTAACAGAGATCTTGAATTAAGATCAGGAAATAG 120  
 QY 821 agaaattgagaatgaactgtgaatgaagaacaaacactgtgaacagagagcagagaag 880  
 DB 121 AGAAATTGAGAAATGAACTGGATGAAGCAAAACACTTGGAAACAGAGCAGCAGAGAG 180  
 QY 881 cagccctggcagagagagagatgctgtgactaacagaaactctggcgcaactctgaacc 940  
 DB 181 CAGCCCTGGCCAGAGAGAGATGCTTGAGACTTAACAGAACTCTGGCGAACTTGAAGCC 240  
 QY 941 aactgcacctcaccagatctgaataagctcaactcagctcaagaaaaaaggtatatactatg 1000  
 DB 241 AACGCAACCTCACAGATCTGAATAGTCAATCTCAGTCAAGAAAAAGGATATCATATG 300  
 QY 1001 ataaattgggaagattcagagaagaatgaagaattggaggaacagtgtgtccagcatg 1060  
 DB 301 ATAAATTGGGAAATTTACAGAGAAATGAAGAAATGGAGAAACAGTGTCTCCAGCATG 360  
 QY 1061 ggaagatgacatgagaatgaagaagaagcctaagcagctgataaagcacaagcagaagcca 1120  
 DB 361 GGAGAGATCATGACACATGAAAGAGGCTTAAGGCACTGGATTAAGCCAGCCAGCCAGCCA 420  
 QY 1121 cagccacagcagctgtgtcagctcctcagcagaagcagaacacagctctctctgtgaagggcaga 1180  
 DB 421 CAGCCCAAGCAGCTGTGTGAGCTCTCTCAGCAAGCAGAAACACACTTCTCTGTGAGAGCGAGA 480  
 QY 1181 gctgtctggaagaggtggaaccgctgtcgagcacaagttaccagcatgcacaaatctgatt 1240  
 DB 481 GCCGTGTGGGAAGAGTGGACCGCTGGCGGACCCAGTTAACCCAGCATGCGCAAAATCTATT 540  
 QY 1241 gctgacctgatagtgaagaagatgaataatgaatgaataaagaagatattatcatctct 1300  
 DB 541 GCTGACCTGTGGTGAACAGAGTGAATTAATGAATTAACAAAGAGATATTATTCATTCATCT 600  
 QY 1301 ggtttagacttaatatgccaacaagcaccacagactctccaggggtgacacgcctcagcc 1360

DB 601 GGTTAGACTTAATATGACACAGCAGACAGACCTTCCAGGGGTGACACGCGCTCAGCC 660  
 QY 1361 tgcagtggtggtgtgtctctcatcaacgagggcgctgtcccgacagcagtggtggtgagc 1420  
 DB 661 TGCAGTGGGGTGTGCTCTCATCAACGGGGCGCTGTCCCGCACAGTGGGTGGAGC 720  
 QY 1421 tggagctgactcagctgagcagacactcctgtgtatgttttcagaaatggtcgaagt 1480  
 DB 721 TGGAGTGTGACTCTAGCTGAGCAGACATCTCTGTATGTTTTCAGAAATGCTTGAAGTT 780  
 QY 1481 atgtgttaaatctgtcctcatcgtatgtcgtatgatatcatatgatttccaataatgaac 1540  
 DB 781 ATGAGTTAAATCTGCAATCTGATCTGATGATATGATGATGATTTTCAATTAATGAAC 840  
 QY 1541 tttttaagaaga 1552  
 DB 841 TTTTAAAGAAA 852

## RESULT 4

VB9187  
 ID VB9187 standard; cDNA; 310 BP.  
 AC VB9187;  
 DT 15-FEB-1999 (first entry)  
 DE EST clone CH699.  
 KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;  
 KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;  
 KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;  
 KW gene therapy; ss.  
 OS Homo sapiens.  
 PN WO9845435-A2.  
 PD 15-OCT-1998.  
 PF 10-APR-1998; U06955.  
 PR 10-APR-1997; US-838821.  
 PA (GEMV ) GENETICS INST INC.  
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,  
 PI Racine JA, Spaulding V, Treacy M;  
 DR WPI; 99-070077/06.  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries.  
 PS Claim 1; Page 138; 61pp; English.  
 CC The present sequence represents a human expressed sequence tag (EST).  
 CC The polynucleotide, which is a secreted EST, and the encoded protein  
 CC are predicted to have useful biological activities which would make  
 CC them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is  
 CC given. Suggested activities include nutritional activity, immune  
 CC stimulating or suppressing activity, haematopoiesis regulating  
 CC activity, tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, haemostatic and thrombolytic  
 CC activity, receptor/ligand activity, anti-inflammatory activity,  
 CC cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The polynucleotide may also be useful for gene therapy.  
 SQ Sequence 310 BP; 121 A; 48 C; 87 G; 51 T;

Query Match 18.0%; Score 279.4; DB 1; Length 310;  
 Best Local Similarity 95.3%; Pred. No. 1e-65;  
 Matches 286; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 606 ggaagaagttacaaagaagaatttaagctatataatcaactggaagaatttcaaacca 665  
 DB 7 GGCCAAAGAGNCGTAGGGAAGGATTTCACTATTAACTGAGGAATTCAAACCA 66  
 QY 666 gctgtctctcgggaatgatatcacaaagtggtggaagaatgctatcagctgaa 725  
 DB 67 GCTGTCTCTTCGGAAATGATATGATATCAAAAGCTGTGAGAAATGGCGTATCATCGTGA 126  
 QY 726 taaacaacatgtgagaagatgagcagaagaagagcagacagagagttcagagcaaac 785  
 DB 127 TAAACCAACATGAGAGAGATGAGCGAGAAAGGAGACAGAGAGTTTCAGAGCAAAAC 186

OY 786 taacaggatcttgaatataatgaagaataatgaagaataatgaagaatga 845  
|||||  
DB 187 TAACAGGATTTTGAATTAAGATCAGAAATAGAAATATGAAATGAACTGATCA 246  
OY 846 aagcaaacacacacttggacagagacagagagagagagagagagagagag 905  
|||||  
DB 247 AAGCAACAAACACTTGGACAGAGACAGACAGAAAGCANCCTGGCCAGAGAGAGATTCT 306

## RESULT 5

T24815/c  
ID T24815 standard; cDNA to mRNA; 317 BP.  
AC T24815:  
DT 30-OCT-1996 (first entry)  
DE Human gene signature HDMS06902.  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
PN W0514772-A1.  
PD 01-JUN-1995.  
PF 11-NOV-1994; J01916.  
PR 12-NOV-1993; JP-355504.  
PA (MATSU) MATSUBARA K.  
PA (OKUBO) OKUBO K.  
PI Matsubara K, Okubo K;  
DR WPI: 95-206931/27.  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
PS Claim 1; Page 1702; 2245bp; Japanese.  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in T19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(7) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
SQ Sequence 317 BP; 92 A; 59 C; 68 G; 82 T;

Query Match 7.4%; Score 114.4; DB 1; Length 317;  
Best Local Similarity 90.3%; Pred. No. 1.8e-21;  
Matches 140; Conservative 0; Mismatches 13; Indels 2; Gaps 2;

OY 1397 ccccgacagcaatcggtcggtgagc-tggagctgaactgaactgaactgaactga 1454  
|||||  
DB 288 CCCGCGACGCACTCGGCTGGAGCAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAG 229  
OY 1455 tatgtttcaagaatgagctgaagatgtgttaaatgtgtcattgtatcagtgat 1514  
|||||  
DB 228 NATGCTTNCAGAAATGCGCTTGAAGCTAATGTGTTTAAACTGCACATTCGATGCTA 169  
OY 1515 atacatalgatttcaataaatgaacttttaag 1549  
|||||  
DB 168 ATACATATGATTTTCAATTAATGAAGACTTTTAAAG 134

RESULT 6  
X51883  
ID X51883 standard; DNA; 432 BP.  
AC X51883:  
DT 22-JUN-1999 (first entry)

DE Human secreted protein 5' EST SEQ ID NO: 97.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;  
KW forensic; gene therapy; chromosome mapping; signal peptide;  
KW upstream regulatory sequence; cytokine activity; cell proliferation;  
KW differentiation; haematopoiesis regulation; tissue growth regulation;  
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
OS Homo sapiens.  
PN W0906552-A2.  
PD 11-FEB-1999.  
PF 31-JUL-1998; IBI236.  
PR 01-AUG-1997; US-905223.  
PA (GEST) GENSET.  
PI Duclert A, Dumas Mline Edwards J, Lacroix B;  
DR WPI: 95-153782/13.  
DR P-PSDB: Y13083.  
PT New isolated brain-derived nucleic acids - used to develop products  
PT which may have cytokine, immune, regulatory, haematopoiesis  
PT regulating, anti-inflammatory or tumour inhibition activity  
PS Claim 1; Page 246-247; 577bp; English.  
CC X51787 to X52019 represent 5' expressed sequence tags (ESTs) for human  
CC secreted proteins, and encode the proteins given in Y12987 to Y13219,  
CC respectively. The proteins given represent the signal peptide and an  
CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
CC can be used for producing secreted human gene products. They can also  
CC be used to develop products for diagnosis and therapy. The proteins  
CC obtained may have cytokine activity, cell proliferation/differentiation  
CC activity, haematopoiesis regulating activity, tissue growth regulating  
CC activity, reproductive hormone regulating activity, chemotactic/  
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
CC or other activities. The products can be used in forensic, gene therapy  
CC and chromosome mapping procedures. The sequences can also be used for  
CC obtaining corresponding promoter sequences. The nucleic acids encoding  
CC the signal peptide can be used for directing extracellular secretion of  
CC a polypeptide or the insertion of a polypeptide into a membrane, or  
CC importing a polypeptide into a cell  
SQ Sequence 432 BP; 144 A; 68 C; 93 G; 125 T;

Query Match 7.3%; Score 113.8; DB 1; Length 432;  
Best Local Similarity 97.2%; Pred. No. 3e-21;  
Matches 137; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

OY 556 tcaagaatgtcatcttctcagaatattcccaactggagccagtggaagaagt 615  
|||||  
DB 292 TTAACAGATGTGATCTTGTCTCAGAAATATGCCCACACTGGAGGCCAGTGAAAGGTT 351  
OY 616 acaagagaaagattcgaactataatcaactgagag-aattcaaacgcaactgcttc 674  
|||||  
DB 352 ACNAAAGAAAGATTTTCACCTAATTAATCAACGTGAGAAATTCNAAAGCAGCTGCTTC 411  
OY 675 tcggg-aatgagtgcacaa 694  
|||||  
DB 412 TCGGGAATAATGATGTCACAA 432

RESULT 7  
O87587  
ID O87587 standard; DNA; 1686 BP.  
AC O87587:  
DT 19-DEC-1995 (first entry)  
DE DNA encoding Leucocytozoan protozoa structural protein epitope.  
KW leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;  
KW leucocytozoanosis; treatment; ss.  
OS Leucocytozoan protozoa sp.  
PN J07089395-A.  
PD 04-APR-1995.  
PF 10-SEP-1993; 226078.  
PR (DOBU-) DOBUTSUYO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.  
PA (NISS-) NISSEIKEN KK.  
DR WPI: 95-167252/22.



1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 84



PA (ARNO/) ARNOT D E.  
 PI Arnot DE, Enea V, Nussenzwei RS, Nussenzweig V;  
 DR WPI: 87-037250/05.  
 DR P-PSDB: P70708.  
 PT New Plasmodium vivax circumsporozoite protein - and synthetic  
 PT peptide(s) cont. its dominant epitope, useful in anti-malarial  
 PT vaccines  
 PS Disclosure; fig. 3; 32pp; English.  
 CC The circumsporozoite protein of *P. vivax* sporozoites encoded by  
 CC this gene is useful in the construction of a central domain of P70704  
 CC vaccine. The sequence consists of a central domain of P70704  
 CC repeated 19 times plus N- and C-terminal regions (practically  
 CC homologous with the corresponding domains of *P. cynomolgi* and *P.*  
 CC knowlesi. See also P70704-07, P70709 and N71065.  
 SQ Sequence 1908 BP; 677 A; 353 C; 470 G; 408 T;

Query Match 3.0%; Score 46; DB 1; Length 1908;  
 Best Local Similarity 44.8%; Pred. No. 0.0093;  
 Matches 261; Conservative 0; Mismatches 315; Indels 6; Gaps 2;

QY 601 caggtggaagaaggttacaagaagaatttcagctattatcaactcggaggaattcaa 660  
 DB 401 CAGAACCAAAATCCACGTAAAATAGCTGAACAACCGAGACAGACAGATGGAC 460  
 QY 661 agccagctgcttcgggaatgtatgcacaagaagtgtgtggagaatcgcctatca 720  
 DB 461 AGCCAGCAGGAGAGAGAGACAGATGAGACGCCAGAGTATGAGAGATGACACAG 520  
 QY 721 ctgaataaacaacatggaga---agatgaggcagaagaaggacagagttcaga 777  
 DB 521 CAGAGAGATAGAGCAGCTGGACACAGCAGAGATAGAGATGAGATGAGATGAGACA 580  
 QY 778 gcaaaaactaacagaggtcttgaataataagatcagaataaaggaaattgagatatga 837  
 DB 581 GAGCAGATGGACACCCAGCAGAGACAGACAGATGATGACAACCGAGAGACAGAG 640  
 QY 838 ctgagtgaagaacacaacacttggaaacagagacagaagagcccttggcagagag 897  
 DB 641 CAGATGGACAAACCGAGGTGATGAGAGAGCTGGAACAACCGAGGTGATGAGACAGTG 700  
 QY 898 gagtgcttgagactaacaagaactcgtggcgaactctgagcaacaaactgcactcaca 957  
 DB 701 GACAAACGACGAGAGATAGAGCAGATGAGACAGCCAGAGATAGAGATGAGATGAGAGC 760  
 QY 958 tctgaataagctcaactcagtcagtcagaagaagaaggtatcatatgataattgysaagta 1017  
 DB 761 CAGCAGAGATAGAGCAGATGAGACAGCCAGAGAGATAGAGCAGCTGGACAGCCAGCAG 820  
 QY 1018 cagaagaagaatgagaatgtggagaaacagtgtgtccagcattggagagatcagagag 1077  
 DB 821 GAGATAGAGCAGATGAGACAGCCAGAGAGATAGAGCAGCTGGACAGCCAGCAGAGA---G 877  
 QY 1078 atgaagaagaagcttaagcagcttggatagaagcagcagccacagccagcagctgtgt 1137  
 DB 878 ATAGAGCAGCTGAGACAGCCAGAGAGATAGAGCAGCTGGACAGCCAGAGAGATAGAG 937  
 QY 1138 cagctcctcagcagcagaacccagcttctcctggagagagcag 1179  
 DB 938 CAGCTGAGACGCCAGAGAAATGTGACAGTGGACAGGAG 979

## RESULT 12

V21511  
 ID V21511 standard; DNA; 6755 BP.  
 AC V21511;  
 DT 17-AUG-1998 (first entry)  
 DE Staphylococcal bacteriocin BacRI operon.  
 KW BacRI; bacteriocin; antimicrobial; antibacterial; antibiotic;  
 KW Moraxella bovis; infectious bovine keratoconjunctivitis; cancer;  
 therapy; ds.  
 OS Staphylococcus aureus strain UT0007 (ATCC 55800).  
 PN MO9812319-AL.

PD 26-MAR-1998.  
 PE 18-SEP-1997; U16758.  
 PE 17-SEP-1997; US-931999.  
 PR 19-SEP-1996; US-710561.  
 PA (UNIV ) UNIV KANSAS STATE RES FOUND.  
 PI Crupper SS, Iandolo JJ;  
 DR WPI: 98-230316/20.  
 PT Therapeutic proteinaceous substances from *Staphylococcus aureus* -  
 PT useful to inhibit growth of wide range of prokaryotic or eukaryotic  
 PT cells, e.g. *Moraxella bovis* causing infectious bovine  
 PT keratoconjunctivitis  
 PS Claim 2; Page 19-23; 38pp; English.  
 CC This polynucleotide comprises the bacteriocin BacRI operon of  
 CC *Staphylococcus aureus* UT0007. The sequence of the BacRI operon  
 CC was determined by N-terminal sequencing of purified BacRI peptide  
 CC (see W54171), with back-translation and plasmid analysis. The  
 CC BacRI operon includes the BacRI gene (see V21510), a homologue of  
 CC the *cym* gene of the cytolytic operon of *Enterococcus faecalis*  
 CC whose function is involved in the maturation of pre-cytolysin,  
 CC an ATP-transporter gene, *bio1* and *bio2* genes related to  
 CC lactococcal biosynthesis and modification, and a gene involved in  
 CC immunity function. BacRI peptides can be produced by construction  
 CC of an expression vector containing an oligonucleotide or operon  
 CC coding for BacRI, and use of the vector to transform host cells for  
 CC BacRI expression. The entire BacRI operon has been cloned into  
 CC plasmid PUB110, and *Bacillus subtilis* transformants secreted the  
 CC recombinant BacRI peptide into the medium. Bacteriocin BacRI is  
 CC active against many Gram-positive and Gram-negative organisms such  
 CC as *Bordetella bronchiseptica*, *Pasteurella multocida* and  
 CC *Staphylococcus aureus*; *Moraxella bovis*, causing infectious bovine  
 CC keratoconjunctivitis, is especially sensitive. BacRI can also be  
 CC used as an anti-cancer agent.  
 SQ Sequence 6755 BP; 3903 A; 1252 C; 1600 G; 0 U;

Query Match 2.9%; Score 44.8; DB 1; Length 6755;  
 Best Local Similarity 42.7%; Pred. No. 0.037;  
 Matches 286; Conservative 0; Mismatches 382; Indels 2; Gaps 1;

QY 461 gacttgaagaagaacttgatcctcagcaagaagaaggccttgatgaagaatgtgtga 520  
 DB 3172 GAACGAAACAAACGCGCAACAGCCAAACAAACGCGCA---GGGAAAGACAGACGA 3229  
 QY 521 aagaagaatacaagaagaaggtgacatggtgataaagtgtgattgtctcaga 580  
 DB 3230 AAGGGGCAAGCAGAGGCAACGAAACAAACGAAACAAACGCGCAACCAACCAAGAGG 3289  
 QY 581 atattcccaacttggagcccggttgaaagaaggttacaagaagaagattcagctatta 640  
 DB 3290 AGAAGCACCCACGAAAGCGCCAGCAGAAAGACGAAACGAAAGAAACCAACCAACGAA 3349  
 QY 641 atcaactggaggaatccaagcagctgtctctcgggaatgtgatatgaagaagttgt 700  
 DB 3350 GCCACGAGCAGGAAAGCAAAAGAGAGACAGGAAAGAGAAAGGCGCAACCAAGAGA 3409  
 QY 701 gtgagaagaatgcgtatcagctgataaacaacatcaggaagaagtgtgagcagaaga 760  
 DB 3410 AAAAAAAGGAAAGAAACAAACGAAAGAAAGAGGGGAGAGAGAAAGAAAGAGAGACA 3469  
 QY 761 agcagagaagttcagaagaactaacaagagatcttgaatlaaagatcagaagaatag 820  
 DB 3470 AAAAAACCGAGCAGAGAAAGAAAGAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAGAA 3529  
 QY 821 agaatgtgaataagaactgtgataagaagcaacaacacttggagaacggagcagagaag 880  
 DB 3530 ACAAGACGAGAGAGAAAAAAGAAAGCAAGCAGACAGAAAGAAAGACCGGAGAAACAC 3589  
 QY 881 cagccctggcagagagagtgcttgactaacaagaactgctgggaggaatcgaagcacc 940  
 DB 3590 GAAAAAAGCAAGAGAGAGAAAAAAGAAAGCAAGCAGCAACCGCAAGCGACCAAAAA 3649  
 QY 941 aactgcaactcaccagatctgaaatagctcaactcagttcaagaagaagttatcataatg 1000



```
Db 3650 AAAAAAGCCGAAACCAAGAGAAAAAACCAAAAGAGAAAGAAAAAAGAAAAAGCGAG 3709
Qy 1001 ataattgggaagttaacagagaagaatgaatggaggaatggaggaatggtccgcatg 1060
Db 3710 CAAAAAGCGGAGAAAAAGGAGGAAACAAAGAAAAAACCAAAAAAGGAAAAAGCAAAA 3769
Qy 1061 ggaagtcacagagacgaatgaagcaaaagctgaagcagctgataagcagcagcgcca 1120
Db 3770 AAGCGGAGACAGCCACCAAAAAAAGGACAGAGAAAAAGCAAAAGAAAAACCAACAA 3829
Qy 1121 cagccagca 1130
Db 3830 AAGAGAAAGCA 3839

RESULT 13
ID T05868 standard; DNA; 3399 BP.
AC T05868;
DE 14-AUG-1996 (first entry)
KW Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
KW Chicken leucocytozoan; Immunogen; recombinant vaccine; protection;
OS Chicken leucocytozoan.
FH Key Location/Qualifiers
FT cds 1 3399
FT /*tag= a
FT misc.feature 1150..3218
FT /note= b
FT /note= "fragment referred to in the claims, for
FT use as insert in a recombinant vaccine
FT against chicken leucocytozoan disease"

J07284392.A.
PD 31-OCT-1995.
PR 19-APR-1994; 080643.
PR 19-APR-1994; JP-080643.
PA (DOBU-) DOBUTSUJO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA (KITA) KITASATO KENKYUSHO SH.
DR WPI: 96-006311/01.
DR P-PSDB: K97866.
PT Chicken leucocytozoan immunogenic protein - used in a recombinant
PT vaccine against chicken leucocytozoan disease
PS Claim 6; Page 6-9; 35pp; Japanese.
CC T05868 encodes a chicken leucocytozoan immunogenic protein, this DNA
CC or a fragment of it can be used in a recombinant vaccine to immunise
CC against chicken leucocytozoan disease. The DNA is used in a vector
CC and operatively linked to an expression regulatory sequence as in
CC standard practice.
SQ Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T;
```

```
Db 2642 TAACATCATGAGAGAGAAAAAGTAACATCATGAGAGAGAAAAAGTAATACATGAAG 2701
Qy 719 agctgaataaaaccacatggaagatgagcgagaaaggacagagaatcagag 778
Db 2702 AAGAAAGAAAAAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2761
Qy 779 caaaactcaacggagctctgaaatgaatcagagaatagaagaatggaatagaagac 838
Db 2762 AAGAAAGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2821
Qy 839 tggatgaaagcaaacacacttggaacaggaagcagcagcagcagcagcagcagcag 898
Db 2822 AAGAAAGAAAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2881
Qy 899 agtccttgactaacagaaactgctggcggaatctgagcacaactgcaactcaccagat 958
Db 2882 AAGACATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2941
Qy 959 ctgaatagctcaactcagtcaggaagaaaggatatacatatgataatggaagttac 1018
Db 2942 AAGAAAGAAAGTAACATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3001
Qy 1019 agagaagaatgaagaattggaagacagtgctcagcatggaagatgacatgagacga 1078
Db 3002 TAACATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3061
Qy 1079 tgaagcaaaagcgaagcagctggaatgaagcacagccagccagccagcagctgtgc 1138
Db 3062 AAGAAAGAAAAATGTAACATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3121
Qy 1139 agctctcagcaagagacagcagctctcctctggaagcagagcagctgtcgggaagcg 1198
Db 3122 AAAAAAGTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3181
Qy 1199 accggtcggaagccagttaccacagcatgcccacaatctgat 1239
Db 3182 AAGGGGGGGGGCCATGTCATCCGGAGTCCTCGAATTCGAT 3222

RESULT 14
ID T46187 standard; DNA; 3586 BP.
AC T46187.
DE 14-APR-1997 (first entry)
DE DNA encoding cell protein homologous to the SV40 large T antigen.
KW Cellular homologue of the SV40 large T antigen; CHLA-1;
KW natural killer lymphic associated protein; tumour suppressor;
KW cancer; viral infection; therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT exon 1..3069
FT /*tag= a
FT /codon.start= 48..50
FT intron 3070..3177
FT /*tag= b
FT exon 3178..3586
FT /*tag= c
FT /note= "in-frame stop codon at 3285-3287"

W09638540-A1.
PD 05-DEC-1996.
PR 30-APR-1996; U07691.
PR 30-MAY-1995; US-453701.
PA (REBE-) RES DEV FOUND.
PI Fung YK;
DR WPI: 97-034359/03.
DR P-PSDB: W01535.
PT DNA encoding cellular protein homologous with SV40 large T antigen -
PT useful for regulating initiation of DNA replication, e.g. for
PT treating tumors and viral infection
PS Example 10; Fig 11A; 59pp; English.
CC A DNA sequence (T46187) encodes a cellular homologue (W01535)
CC of the SV40 large T antigen (CHLA), characterised as a regulator of
CC initiation of DNA replication in cells that is useful e.g. in
```



```
GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on:      May 19, 2000, 22:33:29 ; Search time 3256.04 Seconds
              (without alignments)
              1615.792 Million cell updates/sec

Title:      US-09-502-945-3
Perfect score: 1298
Sequence:    1 ggcctgcgaatgactcgcga.....gaagaagaagaagaaaaa 1298

Scoring table:  IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:      4857316 segs, 2026611650 residues

Total number of hits satisfying chosen parameters:      9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
                  Listing first 45 summaries

Database :
EST:*
1:  em_est1:*
2:  em_est2:*
3:  em_est3:*
4:  em_est4:*
5:  em_est5:*
6:  em_est6:*
7:  em_est7:*
8:  em_est8:*
9:  em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
```

```
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_est1:*
83: gb_est2:*
84: gb_est3:*
85: gb_est4:*
86: em_est1:*
87: em_est2:*
88: em_est3:*
89: em_est4:*
90: gb_est5:*
91: gb_est6:*
92: gb_est7:*
93: gb_est8:*
94: gb_est9:*
95: em_est5:*
96: em_est6:*
97: em_est7:*
98: em_est8:*
99: em_est9:*
100: em_est10:*
101: em_est11:*
102: gb_est10:*
103: gb_est11:*
104: em_est12:*
105: gb_est12:*
106: gb_est13:*
107: gb_est14:*
108: gb_est15:*
109: gb_est16:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| No.  | Score | Match | Length | DB  | ID        | Description         |
|------|-------|-------|--------|-----|-----------|---------------------|
| C 1  | 595   | 45.8  | 643    | 51  | A1733329  | A1733329 op20c03.x  |
| C 2  | 415   | 32.0  | 507    | 39  | AA824593  | AA824593 oc83d10.s  |
| C 3  | 358   | 27.6  | 323    | 29  | AA151187  | AA151187 zo03c11.r  |
| C 4  | 355.4 | 27.4  | 500    | 48  | A1614794  | A1614794 mm43c03.y  |
| C 5  | 319   | 24.6  | 361    | 20  | T16360    | T16360 NIB1193 Not  |
| C 6  | 315.2 | 24.3  | 388    | 94  | AO028402  | AO028402 C1F-HSP-2  |
| C 7  | 308.6 | 23.8  | 472    | 22  | R54592    | R54592 yg81h10.s1   |
| C 8  | 297.4 | 22.9  | 327    | 40  | AA954862  | AA954862 op20c03.s  |
| C 9  | 294.8 | 22.7  | 433    | 28  | AA065677  | AA065677 mm43c03.r  |
| C 10 | 264   | 20.3  | 268    | 32  | AA373594  | AA373594 EST85631   |
| C 11 | 264   | 20.3  | 330    | 79  | AW291601  | AW291601 UI-H-B12-  |
| C 12 | 222.6 | 17.1  | 449    | 22  | R42009    | R42009 yg05b04.s1   |
| C 13 | 214.8 | 16.5  | 521    | 28  | AA100660  | AA100660 z190a05.r  |
| C 14 | 208.2 | 16.0  | 987    | 82  | CNS00057E | AL060767 Drosophill |
| C 15 | 197.2 | 15.2  | 273    | 59  | AV118269  | AV118269 AV118269   |
| C 16 | 195   | 15.0  | 1059   | 82  | CNS00022B | AL097133 Drosophill |
| C 17 | 191.8 | 14.8  | 952    | 91  | AO897460  | AO897460 HS_-3134_A |
| C 18 | 190.2 | 14.7  | 756    | 82  | CNS0118D  | AL099943 Drosophill |
| C 19 | 189.4 | 14.6  | 732    | 103 | AO257374  | AO257374 nbxb0018K  |
| C 20 | 187.8 | 14.5  | 1042   | 83  | CNS0148K  | AL103838 Drosophill |
| C 21 | 186.8 | 14.4  | 982    | 82  | CNS0073S  | AL066780 Drosophill |
| C 22 | 185.4 | 14.3  | 283    | 59  | AV119714  | AV119714 AV119714   |
| C 23 | 185.4 | 14.3  | 870    | 105 | AO330286  | AO330286 nbxb0046T  |
| C 24 | 185.4 | 14.3  | 987    | 82  | CNS00418  | AL065537 Drosophill |
| C 25 | 184.8 | 14.2  | 710    | 80  | AW349204  | AW349204 GM210004A  |
| C 26 | 184.2 | 14.2  | 817    | 82  | CNS0009FM | AL055514 Drosophill |
| C 27 | 183.6 | 14.1  | 736    | 82  | CNS0009DE | AL055636 Drosophill |
| C 28 | 182.6 | 14.1  | 700    | 62  | A1906328  | PM-B1107-           |
| C 29 | 182   | 14.0  | 255    | 48  | A1611494  | A1611494 t165a09.x  |
| C 30 | 182   | 14.0  | 1098   | 90  | AO782501  | AO782501 HS_-3174_A |
| C 31 | 181.8 | 14.0  | 1198   | 84  | B08337    | B08337 t11P9-Sp6.1  |
| C 32 | 181.4 | 14.0  | 830    | 105 | AO291956  | AO291956 nbxb0040O  |
| C 33 | 181.2 | 14.0  | 839    | 82  | CNS0122R  | AL101037 Drosophill |
| C 34 | 181.2 | 14.0  | 939    | 82  | CNS000CBZ | AL058374 Drosophill |
| C 35 | 181   | 13.9  | 506    | 82  | CNS0009K4 | AL053777 Drosophill |
| C 36 | 180.6 | 13.9  | 774    | 107 | AO501198  | AO501198 y25G12.mT  |
| C 37 | 180.4 | 13.9  | 1152   | 90  | AO779752  | AO779752 HS_-5572_B |
| C 38 | 179.6 | 13.8  | 507    | 27  | W82081    | W82081 me96h06.r1   |
| C 39 | 179.2 | 13.8  | 952    | 83  | CNS014BFE | AL103941 Drosophill |
| C 40 | 179   | 13.8  | 983    | 105 | AO330267  | AO330267 nbxb0046H  |
| C 41 | 178.8 | 13.8  | 1036   | 82  | CNS000599 | AL057797 Drosophill |
| C 42 | 178.6 | 13.8  | 864    | 70  | AW155256  | AW155256 mg1e0002P  |
| C 43 | 178.6 | 13.8  | 968    | 85  | AO687544  | AO687544 nbxb00751  |
| C 44 | 178.2 | 13.7  | 1101   | 83  | CNS0153V  | AL104965 Drosophill |
| C 45 | 177.8 | 13.7  | 707    | 94  | AO051636  | AO051636 nbxb0002C  |

## ALIGNMENTS

| RESULT     | 1  |
|------------|--|
| LOCUS      | AI733329   |
| DEFINITION | AI733329 643 bp mRNA EST 14-JUN-1999   |
| ACCESSION  | op20c03.x5 NC1:CGAP.C012 Homo sapiens cDNA clone IMAGE:1576228 3'  |
| VERSION    | similar to TR:092355 Q92555 MYELOBLAST KIAA0266. ;, mRNA sequence  |
| KEYWORDS   | AI733329.1 GI:5054442  |
| SOURCE     | EST.   |
| ORGANISM   | human.   |
| REFERENCE  | Homo sapiens   |
| AUTHORS    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  |
| TITLE      | Euthelia; Primates; Catarrhini; Hominiidae; Homo.  |
| JOURNAL    | 1 (bases 1 to 643)   |
| COMMENT    | NC1/MIDR-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .<br>National Cancer Institute / National Institute of Dental Research<br>Cancer Genome Anatomy Project (CGAP), Tumor Gene Index<br>Unpublished (1997)<br>On May 9, 1996 this sequence version replaced gi:1132809.<br>Contact: Robert Strausberg, Ph.D.<br>Tel.: (301) 496-1550<br>Email: Robert.Strausberg@nih.gov |

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-GAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILMUT at: [www-bio.1ml.gov/dbip/image/image.html](http://www-bio.1ml.gov/dbip/image/image.html)

This read is a RESOURCE of a previously sequenced human clone  
 This original clone citation: National Cancer Institute, Cancer Genome  
 Anatomy Project (CCAP), Tumor Gene Index  
 This read has been verified (found to hit its original self in the  
 correct orientation)  
 Seq primer: -40UP from GIBCO  
 High quality sequence stop: 455.

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| source   | 1. .643             |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1576228"
/clone_lib="NCI-CCAP_C012"
/sex="mixed"
/lissu_type="colon tumor"
/lab_host="SOLR (kanamycin resistant)"
/ncore="Organ: colon; Vector: Bluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'
GAATTCGCGACGAC 3' 3' adaptor sequence: 5'
CTCAGCTTTTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."
BASE COUNT      195 a      137 c      172 g      132 t      7 others
ORIGIN

```

| Query Match           | 45.8%  | Score 595         | DB 51    | Length 643 |
|-----------------------|--|-------------------|----------|------------|
| Best Local Similarity | 97.0%  | Pred. No.1.8e-101 |          |            |
| Matches 620           | Conservative 0   | Mismatches 17     | Indels 2 | Gaps 2     |
| QY 1                  | ggtctgtgaataatgaatctgtgaacccggtcttgcagagagcccttcgtgctttagccmcaagga | 60                |          |            |
| Db 4                  | GGCTCTGAATAATGACTGGGAACCGGCTTCACAGAGACCTTGGGCTTTGAGCAACAGCA        | 63                |          |            |
| QY 61                 | agaactagcggatcttgcacaaagaactacctcttgaagtgaagtgaagaatgagggggacaa    | 120               |          |            |
| Db 64                 | AGAACTACCGGATTTGGCCAAAAGACTTACTCTTGTGAGTGAGAGTGAAGAGAGGGGACAA      | 123               |          |            |
| QY 121                | tgatgagaagaagaagacatcnaaagctctcttgaagaacaatcagttcccttgatggaaga     | 180               |          |            |
| Db 124                | TGATGAGAAGAGAAAGCATCAAAAGCTTCTGGAGAGCAATCACTTCCCTTGATGGAAGAA       | 183               |          |            |
| QY 181                | tagcgggaatctgctgtgaaggtctcgaggtcgtctgaagtgctcagagttcaatgtcag       | 240               |          |            |
| Db 184                | TAGCGGAAATTTGGCTGAGAGGCTTGAGGCTAGTCTGAAGGTTGTCAGAGTTCAATGTGAG      | 243               |          |            |
| QY 241                | ttctgaagaatcaggagaagaagctggtctcttgcagatctgcttgaagcctgttaaaattc     | 300               |          |            |
| Db 244                | TTCTGAAAGATCAGAGAAAGACTGTGCTTCCTTCAGATCTGCTTGAGCCCTGTAAAACTTC      | 303               |          |            |
| QY 301                | atctctcttggccaatgtgaaaaagccaactgtgtagatgcataatcaaaanaaacgttga      | 360               |          |            |
| Db 304                | ATCTCTTTTGGCCATGTGAAAAAGCAATGTGATGAGTCAATCAATAAAGACAACTGGGA        | 363               |          |            |
| QY 361                | gttacctctgaacaagaagaagatgtgaacccgtaaccacagaga-tagcaatcaataaa-      | 419               |          |            |
| Db 364                | GTTACTCTGAACAAGAAGAGATTTGAACGGATCCACAGAGAAGTAGATTCAATPAAAC         | 423               |          |            |
| QY 419                | cgcacaagtcctctccaaatgggacctgtgctctgaagaacccggcagcagcagct           | 478               |          |            |
| Db 424                | CGCACAAGTCTCTCTCAATGGGAGACCTGTCTCTCTGAAGAACCGGCAAGCAGACGACT        | 483               |          |            |
| QY 479                | ggttttccctctgggaagaagagagccagccatctgctccatgtgaacatgttctcagtg       | 538               |          |            |
| Db 484                | GGTTTTCCTCCCTGGGAAGAAGAGACCAAGCCATGCTCTCCATTGAACATGTCTCACTGAG      | 543               |          |            |



```
/db_xref="GDB:4591065"
/db_xref="taxon:9606"
/clone="IMAGE:566612"
/clone_lib="Stratagene colon (#937204)"
/tissue_type="tumor"
/cell_line="T84 carcinoma cell line"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site:1;
ECORI; Site:2; XhoI; Cloned unidirectionally. Primer:
Oligo dT. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGACGACG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTT 3'."
BASE COUNT      123 a      69 c      111 g      88 t      1 others
ORIGIN
```

```
Query Match      27.6%; Score 358; DB 29; Length 392;
Best Local Similarity 98.1%; Pred. No. 2.1e-57;
Matches 369; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
```

```
Qy 1 ggcgcgaatgactgacgacccgcttcgacagagaccccttcgttcgaccanacaga 60
    |||||||
Db 18 GGCCTGCTAAATGACTGCGAACCGCTTGACAGAGCCCTCT-GTTTGGACCAACAGCA 76

Qy 61 agaactagcagattgcccaaaagactacctctgagtgaagagatgaagagagagacaa 120
    |||||||
Db 77 AGAACTAGCGGATTTGCCAAAAGACTACTCTTGAGTAGAGTAGAGTAGAGGGGACAA 136

Qy 121 tgatgagagagaagaacatcnaagctctcgtgaagcaatcagcttccttgatgaaaga 180
    |||||||
Db 137 TGATGAGAGAGAGAAAGCATCAAAAGCTCTGGAAGCATCAATGCCCTTGGTGAAGAA 196

Qy 181 tagcgcgaatgactgacgacccgcttcgacagagaccccttcgttcgaccanacaga 240
    |||||||
Db 197 TAGCGGAAATGCTGAGAGCTGAGGCTGAGCTGGAAGGCTGAGAGTTCAATGTCAG 256

Qy 241 ttctgaagatcagagagaagaagctgcttccttcgacagatcgttcgacgtttaaactc 300
    |||||||
Db 257 TTCTGGAAGATCAGAGAGAAAGCTGCTGCTTGCAGATGCTGTAAGCTGTTAAATCTC 316

Qy 301 atctctctgacactgtaaaagcaactgagtagagatcnaatcaaaagaanaacagtga 360
    |||||||
Db 317 ATCTCTCTTGGCCACTGTGAAGCAACTGATGAGTCAATCAAGAAAGACAGTGA 376

Qy 361 gtacacctgaacaaa 376
    |||||||
Db 377 GTTACCTCTGAAACAA 392
```

```
RESULT 4
LOCUS      A1614794      500 bp      mRNA      EST      21-APR-1999
DEFINITION mm43c03.y1 Stratagene mouse melanoma (#937312). Mus musculus cDNA
            clone IMAGE:524260 5' similar to TR:Q92555 Q92555 MYELOBLAST
            KIA0266.; mRNA sequence.
```

```
ACCESSION  A1614794
VERSION     A1614794.1 GI:4623961
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

REFERENCE 1 (bases 1 to 500)

AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Waller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999

JOURNAL Unpublished (1999)

COMMENT On Jun 5, 1998 this sequence version replaced gi:3187033.  
Contact: Marra M/WashU-NCI Mouse EST Project 1999

Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
This read is a RESQUENCE of a previously sequenced mouse clone  
This read has been verified (found to hit its original self in the  
correct orientation)  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from GIBCO  
High quality sequence stop: 425.

## FEATURES

1. 500  
Location/Qualifiers

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:524260"
/clone_lib="Stratagene mouse melanoma (#937312)"
/tissue_type="melanoma"
/dev_stage="M2 cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site:1;
ECORI; Site:2; XhoI; Cloned unidirectionally. Primer:
Oligo dT. From M2 cells, a highly metastatic derivative of
the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGACGACG
3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
BASE COUNT      167 a      111 c      118 g      101 t      3 others
ORIGIN
```

```
Query Match      27.4%; Score 355.4; DB 48; Length 500;
Best Local Similarity 84.0%; Pred. No. 6.6e-57;
Matches 420; Conservative 0; Mismatches 78; Indels 2; Gaps 2;
```

```
Qy 339 tcanatcaagaagaacagtgagttacctcgtgaacaaagaagatgaacgattccaca 398
    |||||
Db 1 TCAATGCAAAAGACTGCTGGAATTAATCTTCAACAAAAGAGAGTTGAACAGATCCACA 60

Qy 399 gaga-tagcatcaataaac-gcaacagtcctctccaaatggaacctgctcctga 456
    |||||
Db 61 GAGAAAGTGGCATTCAGTAACTTCAACAAACCTCTCCAAATGGGATTCGTGTTGAGA 120

Qy 457 agaaccggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 516
    |||||
Db 121 AGAACCGGAAAGCAGAGAGCTGGCTTTTCTCTGAGAAAGAGCATCTTTTGTCTC 180

Qy 517 ccattgaacatgctcagtgctggaagcaagaactccctctgagcaggaatcttca 576
    |||||
Db 181 CCATGGAACATGTTTCTGTAAGGGAAGCGAAGCACTCCACTGGAACAGAACTTTTGA 240

Qy 577 acctctccataagaacaagcagcagcagcagcagcagcagcagcagcagcagcagcag 636
    |||||
Db 241 ACCTCTCTCATTAACAAACAGCAGCAGACACCTTTACTGATCTCTGTGAAAAAGG 300

Qy 637 cctctcccgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 696
    |||||
Db 301 CTTCTCTTAAGGCGATGAGCTGGAAGAGGCTTAAGATTCGACAGACAGCTTTGAGAGG 360

Qy 697 ctgggctctgactctactactatgancgaagcctcgaagagagagagagagagagag 756
    |||||
Db 361 CCCGAGCTCTGCAATGCTGATGATGAGGCAAGGCTCGAAGAAATGAAATCAAAAGTA 420

Qy 757 aaagatcacaaagctgtaagaagaagcgaagcgaagcgaagcgaagcgaagcgaag 816
    |||||
Db 421 AAAATACCAANAATGTGAAGAAAGGAGGCAAGAAAGCCCTTAAGATTTTGAGC 480

Qy 817 agctgcggaaggttaacca 836
    |||||
Db 481 AGCTGAGAAAGTTGACCA 500
```

[illegible][illegible]

|    |     |  |     |
|----|-----|--|-----|
| Db | 302 | GACCCATATCATCCCTGAGAAACCATCATGACAGATCTGTGTTTTCCCTGGGAATGAG | 361 |
|    |     |  |     |
| QY | 501 | gagccaagccattgctcccatgaaacat                               | 527 |
|    |     |  |     |
| Db | 362 | CAGCCAGCCATTGCTCCCATTTGAACAT                               | 388 |

|            |  |
|------------|--|
| RESULT     | 7  |
| LOCUS      | R54592/c   |
| DEFINITION | R54592 472 bp mRNA EST 18-MAY-1995                         |
| ACCESSION  | Y98110.s1 Soares Infant Brain INIB Homo sapiens cDNA clone |
| VERSION    | R54592 IMAGE:40102.3, mRNA sequence.                       |
| KEYWORDS   | R54592.1 GI:816494   |
| SOURCE     | EST.   |
| ORGANISM   | human.<br>Homo sapiens                                     |

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 472)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman, M., Hultman, M., Kueaba, T., Le, M., Lennon, G., Marra, M.,  
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
Trevaskis, E., Waterston, R., Williamson, A., Woldmann, P., and  
Wilson, R.  
The Washu-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK

**TITLE** The WashU-Merck EST Project  
**JOURNAL** Unpublished (1995)  
**COMMENT** Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watsn.wustl.edu  
Insert Size: 1408  
High quality sequence stops: 393 Source: IMAGE Consortium, LIND  
This clone is available royalty-free through LIND ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert length: 1408 Std Error: 0.00  
Seq primer: Promega -21m13  
High quality sequence stop: 393.

```

FEATURES
SOURCE
location/Qualifiers
1..472
/organism="Homo sapiens"
/db_xref="GDB:412643"
/db_xref="taxon:9606"
/clone="IMAGE:40102"
/clone_1fb="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Latmid BA; Site:1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AACGGAGAAGATTCGGCGCCAGCAAGATTTTTTTTTTTTTTTT 3'];"
(double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia) , digested with Not I and directionally cloned
into the Not I and Hind III sites of the Latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT
83 a 111 c 118 g 157 t 3 others

```

|                      |              |                   |               |            |
|----------------------|--------------|-------------------|---------------|------------|
| Query Match          | 23.8%        | Score 308.6       | DB 22         | Length 472 |
| Best Local Similarly | 86.7%        | Pred. No. 3.3e-48 |               |            |
| Matches 385          | Conservative | 0                 | Mismatches 52 | Indels 7   |
|                      |              |                   |               | Gaps 4     |

Oy 429 cctcccaatgtagaacctgtgcgtcctaagaaccggacagcgacgcgtgtttt--c 468  
||| || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 446 CTCCCAATGGGACCCTTATCATCTCTGAAGAACCGACGACGAGCAGNCTGGTTTTC 387  
  
Oy 487 ccctcgagaaagtgagtcagcca---tgtctcccattgaacatgtg-ctcagtgtgctcg 542

Db 386 CCCTGGGGAGGAGACACCCAGCCCATTTGCTCCCCATTGACATGAGCGCTCACTGGCTGG 327

QY 543 aaggcaagaact-ccccitygacgagaatttccaacctctccataaagaacaagagcc 601

Db 326 AAGGCAGAATCCTCCCTCGAGCAGGAATTTTAACCTCTCCATTAAGACAGAGCC 267

QY 602 atgacagacccttactgacccccitygaaaagcctctccgacatgagcctaga 661

Db 266 AGTGACAGATCTTTACTGTGACTCCCATGAAAAAGGCGCTCTCTCAAGGCATGAGCTGGA 207

QY 662 agagggcaaatgctcgagcagcagagctctcaagaggctcgggtctcagtcctactatga 721

Db 206 AGAGGCAAAAGATGCACCGACAGCAGGCTTCAGAGGGCTCGGGCTCTCAGTCCATCATGA 147

QY 722 gcccaagctctgaagagagagaagaataatcnaaatgtataaagtatcacaaagtctgaaga 781

Db 146 GGCCAAGGCTCGAAAAAGAAAGAAATCAAAATATCAAAAAAATACTATCACAAAGTCGTGAAGAA 87

QY 782 aggaagaagccaaagaagccctaaagaagtttggcgagctcgtcggaagttaatccagctgc 841

Db 86 AGGAAGAGGCCAAGAAACCTTTAAAGATTTGAGCAGCTACAGAAAGTTTAATCCACTGT 27

QY 842 cgactagaagaacgaagaagaaga 865

Db 26 GCATTTGGAAGAAATGAAAAAAA 3

[illegible]

REFERENCE 1 (bases 1 to 327)  
AUTHORS NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151515.

Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: Stratagene, Inc.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrrp/image/image.html](http://www-bio.llnl.gov/bdrrp/image/image.html)  
 Seq primer: -40ml3 fwd. RT from Amersham  
 High quality sequence stop: 311.  
 Location/Qualifiers  
 1..327

```

/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: colon; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer
oligo dT. Pooled colon tumors. 5' adaptor sequence: 5'
GAATTGGCAGCAGG 3' 3' adaptor sequence: 5'

```





Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,  
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,  
He,W.M., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,  
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H.,  
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,  
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,  
Fraser,C.M., and Venter,J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

JOURNAL  
MEDLINE  
COMMENT  
12140200  
On May 9, 1995 this sequence version replaced gi:803089.  
Other ESTs: THC192511  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. 268  
/organism="Homo sapiens"  
/db\_xref="ATCC (lnhost):177963"  
/db\_xref="taxon:9606"  
/clone\_lib="HSCI72 cells 1"  
/cell\_type="fibroblast"  
/cell\_line="HSCI72 (16PDL)"  
/note="Organ: lung; Vector: pBluescript SK-; Site:1:  
EcoRI; Site:2: XhoI"  
BASE COUNT 82 a 45 c 84 g 56 t 1 others  
ORIGIN

Query Match 20.3%; Score 264; DB 32; Length 268;  
Best Local Similarity 98.5%; Pred. No. 6e-40; Indels 0; Gaps 0;  
Matches 264; Conservative 0; Mismatches 4;

QY 5 gctcaaatgactgcgaacccgcttcagagagccttcgtcttgagccanagaagaag 64  
Db 1 GCTCAAAATGACTGCGAACCGCTTGCAGAGAGCCTTCGCTTGAGCCAAACAGAA 60  
QY 65 cttagcgatttgcacaaagactacctcttgagtgagagtgagaatgaggggacaatgat 124  
Db 61 CTAGCGGATTTGGCAAAAGACTACCTCTTGAGTGAAGATGAGGGGACCAATGAT 120  
QY 125 gggagagagaagactcnaaagcttcgtgaagcaatcagtccttcgtatgagaagaatag 184  
Db 121 GGAGAGGAAAGCAATCAAAAGCTTCTGGAACCAATCAATGATCCCTTGATGAAAGAAATAGG 180  
QY 185 cggaaatgctgctgnaagcttcgaagctagtcgtgaaggtgtcagaagttcaatgtaattct 244  
Db 181 CGGAATTTGGCTGAGAGGTCTGAGGCTAGTCTGAAGGTCTCAGACTCAATGCTCAGTNC 240  
QY 245 gaagatcagagaagaagctgtgctctg 272  
Db 241 GAAGGATCAGGAGAAAGCTGCTCTTG 268

RESULT 11  
AW291601 330 bp mRNA EST 16-JAN-2000  
LOCUS AW291601/c  
DEFINITION UI-H-BI2-akg-h-06-0-UI-sl NCL CGAP\_Sub4 Homo sapiens cDNA clone  
IMAGE:2724730 3', mRNA sequence.  
ACCESSION AW291601  
VERSION AW291601.1 GI:6698237  
KEYWORDS EST.  
SOURCE human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
COMMENT  
1 (bases 1 to 330)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jul 9, 1999 this sequence version replaced gi:5434385.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. cDNA library preparation: M.B. Soares Lab Clone distribution:  
NCI-CGAP clone distribution information can be found through the  
I.M.A.G.E. Consortium/LNLN at:  
[www.bio.lnl.gov/dbir/image/image.html](http://www.bio.lnl.gov/dbir/image/image.html)  
Seq primer: M13 Forward  
POLYA=yes.  
Location/Qualifiers  
1. 330  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NCI CGAP Sub4"  
/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pRT3D-Pac (Pharmacia) with a modified  
polylinker; Site:1: Not I; Site:2: Eco RI; The  
NCI-CGAP Sub4 library is a subtracted library derived from  
the NCI-CGAP Sub2 library which is a subtracted library  
derived from the NCI-CGAP Sub1 library, which is a  
subtracted library derived from BI. BI constitutes a  
mixture of 21 normalized or subtracted NCI-CGAP  
libraries: NCI-CGAP-Co4, NCI-CGAP-Pr22, NCI-CGAP-Pr28,  
NCI-CGAP-Co10, NCI-CGAP-Co16, NCI-CGAP-Kid5,  
NCI-CGAP-Kid12, NCI-CGAP-Kid3, NCI-CGAP-Co8, NCI-CGAP-CLL1,  
NCI-CGAP-Lei12, NCI-CGAP-Brn23, NCI-CGAP-Lu5,  
NCI-CGAP-Lu24, NCI-CGAP-Lu19, NCI-CGAP-GC4, NCI-CGAP-GC6,  
NCI-CGAP-Brn25. These 21 libraries were pooled and a  
single-stranded DNA preparation of the resulting mixture  
was used as a tracer in a subtractive hybridization with  
a diver whose composition is detailed below:  
NCI-CGAP Kid3 pool 1: L1AM 3334-3337, 3682-3683,  
3798-3803 (IMAGE Clonoids 1322376-1323911,  
1456008-1456775, 1500552-1502855) NCI-CGAP Kid5 pool 1:  
L1AM 3338-3342, 3722-3725, 3776-3778 (IMAGE Clonoids  
1323912-1325831, 1471368-1472903, 1492104-1493255)  
NCI-CGAP Lu5 pool 1: L1AM 3575-3582, 3851-3854 (IMAGE  
Clonoids 1414920-1417991, 1520904-1522439) NCI-CGAP GC4  
pool 1: L1AM 3164-3167, 3718-3720, 3733-3735 (IMAGE  
Clonoids 1257096-1258631, 1469064-1470983,  
1475592-1476743) NCI-CGAP Pr22 pool 1: L1AM 2457-2459,  
2758-2759, 3062-3068 (IMAGE Clonoids 985608-986759,  
1101192-1101959, 1217928-1220615) NCI-CGAP Co10 pool 1:  
L1AM 2644-2653, 2871-2872 (IMAGE Clonoids 1057416-1061255,  
114584-1145351) Subtraction was performed as previously  
described (Bonaldi, Lennon & Soares (1996): Normalization  
and Subtraction: Two Approaches to Facilitate Gene  
Discovery. Genome Research 6, 791-806.)  
TAG-LTB-NCI CGAP\_Brn23  
TAG-TISSUE=Brain  
TAG\_SEQ=ATATC"

BASE COUNT 59 a 78 c 89 g 104 t  
ORIGIN

Query Match 20.3%; Score 264; DB 79; Length 330;  
Best Local Similarity 87.5%; Pred. No. 6.2e-40;  
Matches 288; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

|    |     |  |     |
|----|-----|--|-----|
| OY | 419 | cgcaagtcctctccaaatgagaccctgtctgtctctgaagaaccggagagcagaagcct  | 478 |
| Db | 330 | CTCACAGGCTCTCTCCCAAAATGGAGCCCTATATCTCTGAAGAACGAGGACGAGCGCT   | 271 |
| OY | 479 | ggttttccctctgagaagaagagagccacatgtctcccatgtgaatgtgtcagtg      | 538 |
| Db | 270 | GGTTTCTCCCTGGGGAGAGACGCCAGCCATTGCTCCATTAAATCGCTCAATGG        | 211 |
| OY | 539 | ctggaagcagaagacccccctgagcagaaatttcaacctctccatagaacaagca      | 598 |
| Db | 210 | CTGGAAGCAGAACTCCCTCGAGACAGAAATTTTAACCTCTCCATAGAACAAGCA       | 151 |
| OY | 599 | gccagtgacagacccttacttgacccccgttgaaaagcctctctccagacatgagcct   | 658 |
| Db | 150 | GCCAGTGACAGATCCTTACTGACTCCOATGAAAGGCGCTCTCCAAAGCATGAGACCT    | 91  |
| OY | 659 | agaagagcagaatgtcgagcagcagacgtctcagaaggctcgagctctgcagtcctacta | 718 |
| Db | 90  | GGAAAGAGCAAAAGATGCAACCGACGAGAGCTTCAAGAGGAAGGGGTCTCTACGATCTTA | 31  |
| OY | 719 | tgangccaagctctgagaagagaagaaaa                                | 747 |
| Db | 30  | TGACCCATGCAAGAAAAAAAAAAAAA                                   | 2   |

|            |   |
|------------|---|
| RESULT     | 12  |
| R42009/c   |   |
| LOCUS      | R42009 449 bp mRNA  |
| DEFINITION | Y905004.s1 Soares infant brain IN1B Homo sapiens cDNA clone |
| IMAGE:     | 31336 3', mRNA sequence.                                    |

ACCESSION R42009  
VERSION R42009.1 GI:819562  
KEYWORDS EST.  
SOURCE human.

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 449)  
REFERENCE  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
AUTHORS

|         |  |
|---------|--|
| TITLE   | The WashU-Merck EST Project                              |
| JOURNAL | Unpublished (1995)                                       |
| COMMENT | On May 8, 1995 this sequence version replaced qi:800233. |

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu  
Insert Size: 1364  
High quality sequence stops: 322 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL ; contact the  
IMAGE Consortium (info@image.jnl.gov) for further information.  
Insert length: 1364 Std Error: 0.00  
Seq primer: Promega -21ml3  
High quality sequence stop: 322.

## FEATURES

### source

```

/organism="Homo sapiens"
/db_xref="GDB:403683"
/db_xref="taxon:9606"
/clone="IMAGE:31336"
/clone_1fb="Scares infant brain IN1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/notes="organ: whole brain; Vector: lambdaB; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dt) primer 15',
LACGTGGAGAAATCCCGCCGCGGACGAGCAATTTTTTTTTTTTTTTT 3']:
```

double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lambda BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

|                       |                 |                    |            |             |
|-----------------------|-----------------|--------------------|------------|-------------|
| Query Match           | 17.1%;          | Score 222.6;       | DB 22;     | Length 449; |
| Best Local Similarity | 85.1%;          | Pred. No. 3.3e-32; |            |             |
| Matches 343;          | Conservative 0; | Mismatches 46;     | Indels 14; | Gaps 8      |

QY 477 ctggttttccctgagaaagaggagcc--agccattgctcccatggaacatg--tgc 531

Db 405 CTGCTTTTCCCTGGGAGAGCAGCCAGCCCATTTGCTCCCCATTGACATGCCGCC 346

QY 532 tcagtgcgtgaag-caagaactccctgag-caggaatttcaacctc--ccat 587

Db 345 TCAGTGGCTGGAAAGGCCAAGAACTCCCTCGGAGCCAGGAATTTTAAACCTCCCTCCAT 286

**OY**    586   aagaacaagcag -ccagtgcacga -ccccttactaccgccctgtgaaagycctcttc 645  
||||||| ||||| ||||| |  
||||||| ||||| ||||| ||

Db 285 AAGAACAAGCAGCCAGTGACAGATCCCTTTACTGACTCCCATGGAAAGGCTCTCTCC 220

QY 646 gagccatgag---cctagaagggcacaagatgcgaagcagcagcttcagagggcttcgg 702

Db 225 CAGCCCATGAGCCCTGGAAAGAGGCCAAGAATGCACCGAGAGCAGAGCTTCAGAGAGGGGCTCGGG 166

003 ctctgcagctctacatcganagcaagctctcgaaagaggaagaaatctnaaagttaaagt/62

DU 103 C1C1GCAGCC1AC1A1GAGGGCCAAAGGC1CGAAAAAGAACAAAT1CAAAGA1AAAAGG1 100

Ov 763 ATCAAGAAATCATAAAGAAAAATTCCTTCAAGAAAGAGGCTAAAAAGATTATTAAGAGGTTC 897

105 ATTCACAAAGCTGTAAGCAAGCAAGCCGAGAAAGCTTTAAAGACAGTGTTCAGCAAGCTTAC 46

[illegible]

Db 45 AGAAGCTTATCCACTGTGGCATTGGAGGAATGGAAAAA 3

RESULT 13  
AA100660

|            |  |        |      |     |             |
|------------|--|--------|------|-----|-------------|
| LOCUS      | AA100660   | 521 bp | mRNA | EST | 31-JUL-1997 |
| DEFINITION | z190a05.r1 Striatagene colon (#937204) Homo sapiens cDNA clone |        |      |     |             |

IMAGE:511856 5', mRNA sequence.  
AA100660  
ACCESSION

```
VERSION AA100660.1  GI:1646978
KEYWORDS EST.
```

|          |              |
|----------|--------------|
| SOURCE   | human.       |
| ORGANISM | Homo sapiens |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

| REFERENCE | 1 (cases 1 to 321)  |
|-----------|---|
| AUTHORS   | Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chizzese, A., Dietrich, N., Edwards, E., Fenucci, A., Gish, E. |

Cuijsse, S., Diecklich, N., Dubuque, I., Fedeliu, A., Gish, W.,  
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
 Mardie, F., Moore, R., Morris, M., Parsons, T., Piffin,

Mullins, E., Moore, D., Morris, M., Parsons, J., Plange, C., Riddick, B., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, Y., Trevaskis, E., Underwood, K., Wohlmann, P., Waterson, R., Wilson, L.

| TITLE  | ISSN      | ONLINE    | WORLDWIDE | WALL      |
|--|-----------|-----------|-----------|-----------|
| Generation and analysis of 280,000 human expressed sequence tags | 1548-7675 | 1548-7675 | 1548-7675 | 1548-7675 |
| and Marra, M.  |           |           |           |           |

JOURNAL  
Genome Res. 6 (9), 807-828 (1996)  
MEDLINE  
97044478

COMMENT  
Contact: Wilson RK  
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800

Fax: 314 286 1810  
Email: [est@atson.wustl.edu](mailto:est@atson.wustl.edu)

This clone is available royalty-free through LBNL ; contact the IMAGE Consortium ([info@image.lbnl.gov](mailto:info@image.lbnl.gov)) for further information





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:30:22 ; Search time 158.18 Seconds  
(without alignments)  
1041.916 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 1298  
Sequence: 1 ggcctgtcgaatgactgcga.....gaagaagaagaagaanaa 1298

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 226296 segs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/6.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/PCFUS.COMB.seq:\*  
7: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | ID                  | Description       |
|------------|-------|-------------|--------|---------------------|-------------------|
| C 1        | 215.4 | 16.6        | 7218   | 1 US-08-232-463-14  | Sequence 14, Appl |
| 2          | 95.4  | 7.3         | 19124  | 4 US-08-487-826B-13 | Sequence 13, Appl |
| 3          | 93.6  | 7.2         | 240    | 1 US-08-628-417-6   | Sequence 6, Appl  |
| 4          | 91.6  | 7.1         | 2447   | 4 US-09-014-969-14  | Sequence 14, Appl |
| 5          | 86    | 6.0         | 291    | 1 US-07-922-723A-7  | Sequence 7, Appl  |
| 6          | 86    | 6.6         | 291    | 1 US-07-799-828C-7  | Sequence 7, Appl  |
| 7          | 86    | 6.6         | 291    | 1 US-08-074-215-7   | Sequence 7, Appl  |
| 8          | 86    | 6.6         | 291    | 1 US-08-480-366-7   | Sequence 7, Appl  |
| 9          | 86    | 6.6         | 291    | 3 US-07-952-277A-7  | Sequence 7, Appl  |
| 10         | 85.8  | 6.6         | 2223   | 3 US-08-257-073-4   | Sequence 4, Appl  |
| 11         | 84    | 6.5         | 494    | 3 US-08-332-766A-22 | Sequence 22, Appl |
| 12         | 82.8  | 6.4         | 376    | 3 US-08-623-906A-18 | Sequence 18, Appl |
| 13         | 82.8  | 6.4         | 43795  | 5 US-08-742-185-101 | Sequence 10, Appl |
| 14         | 82.4  | 6.3         | 4766   | 6 PCT-US93-07261-10 | Sequence 10, Appl |
| 15         | 82.2  | 6.3         | 1859   | 5 US-08-691-563C-46 | Sequence 46, Appl |
| 16         | 81.8  | 6.3         | 446    | 3 US-08-332-766A-26 | Sequence 26, Appl |
| 17         | 80.4  | 6.2         | 454    | 3 US-08-623-906A-6  | Sequence 6, Appl  |
| 18         | 78.6  | 6.1         | 3211   | 4 US-08-574-959A-8  | Sequence 8, Appl  |
| 19         | 78.6  | 6.1         | 3901   | 4 US-08-574-959A-6  | Sequence 6, Appl  |
| 20         | 78.4  | 6.0         | 4698   | 1 US-07-807-043B-5  | Sequence 5, Appl  |
| 21         | 78.4  | 6.0         | 4698   | 1 US-08-299-849B-5  | Sequence 5, Appl  |
| 22         | 78.4  | 6.0         | 4698   | 4 US-08-142-368A-5  | Sequence 5, Appl  |
| 23         | 78.4  | 6.0         | 4698   | 5 US-08-967-727-5   | Sequence 5, Appl  |
| 24         | 78.2  | 6.0         | 285    | 3 US-08-630-822A-85 | Sequence 85, Appl |
| 25         | 78.2  | 6.0         | 285    | 4 US-09-005-069-85  | Sequence 85, Appl |
| 26         | 78    | 6.0         | 1236   | 3 US-08-741-134-5   | Sequence 5, Appl  |
| 27         | 77    | 5.9         | 1598   | 5 US-08-785-420-3   | Sequence 3, Appl  |

|      |      |     |       |                     |                   |
|------|------|-----|-------|---------------------|-------------------|
| 28   | 76.2 | 5.9 | 6243  | 4 US-09-056-075-1   | Sequence 1, Appl  |
| C 29 | 76   | 5.9 | 56516 | 4 US-08-996-306-1   | Sequence 1, Appl  |
| 30   | 74   | 5.7 | 1434  | 4 US-08-903-801-2   | Sequence 2, Appl  |
| 31   | 73.6 | 5.7 | 370   | 3 US-08-332-766A-8  | Sequence 8, Appl  |
| 32   | 73.6 | 5.7 | 3337  | 1 US-08-072-610-1   | Sequence 1, Appl  |
| 33   | 73.6 | 5.7 | 3337  | 3 US-08-719-822B-1  | Sequence 1, Appl  |
| 34   | 73   | 5.6 | 1621  | 4 US-08-933-750C-96 | Sequence 96, Appl |
| 35   | 73   | 5.6 | 5394  | 5 US-08-688-376-1   | Sequence 1, Appl  |
| 36   | 72.4 | 5.6 | 1866  | 5 US-09-173-581-13  | Sequence 13, Appl |
| 37   | 69.4 | 5.3 | 2261  | 1 US-08-272-882D-1  | Sequence 1, Appl  |
| 38   | 68.6 | 5.3 | 1192  | 2 US-08-340-539A-3  | Sequence 3, Appl  |
| 39   | 68.6 | 5.3 | 1192  | 3 US-08-461-592B-3  | Sequence 3, Appl  |
| 40   | 68.6 | 5.3 | 1377  | 4 US-08-810-572A-1  | Sequence 1, Appl  |
| C 41 | 67.6 | 5.2 | 217   | 3 US-08-332-766A-4  | Sequence 4, Appl  |
| C 42 | 67.2 | 5.2 | 421   | 2 US-08-480-784-24  | Sequence 24, Appl |
| C 43 | 67.2 | 5.2 | 421   | 2 US-08-483-553-24  | Sequence 24, Appl |
| C 44 | 67.2 | 5.2 | 421   | 2 US-08-487-002-24  | Sequence 24, Appl |
| C 45 | 67.2 | 5.2 | 421   | 2 US-08-483-554B-24 | Sequence 24, Appl |

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHNEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTZgpt-Fls  
; US-08-232-463-14

[illegible]

RESULT 2  
 US-08-487-826B-13  
 Sequence 13, Application US/08487826B  
 Patent No. 5993827  
 GENERAL INFORMATION:  
 APPLICANT: Sim, Kim L.  
 APPLICANT: Chltnis, Chetan  
 APPLICANT: Miller, Louis H.  
 APPLICANT: Peterson, David S.  
 APPLICANT: Su, Xin-zhaun  
 APPLICANT: Wellens, Thomas E.  
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
 NUMBER OF SEQUENCES: 45  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe Martens Olson & Bear  
 STREET: 620 Newport Center Drive 16th Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: US  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/487,826B  
 FILING DATE: 10-SEP-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelsen, Ned  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH121.001CP1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 19124 base pairs

```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: CDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-487-826B-13

```

[illegible]

RESULT 3  
US-08-628-417-6  
Sequence 6, Application US/08628417  
Patent No. 5627054  
GENERAL INFORMATION:  
APPLICANT: GILLESPIE, DAVID  
TITLE OF INVENTION: COMPETITOR PRIMER ASYMMETRIC  
POLYMERASE CHAIN REACTION  
TITLE OF INVENTION: POLYMERASE CHAIN REACTION  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: U.S. ARMY CHEMICAL AND BIOLOGICAL  
ADDRESSEE: DEFENSE COMMAND  
STREET: OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)  
CITY: ABERDEEN PROVING GROUND  
STATE: MARYLAND  
COUNTRY: USA  
ZIP: 21010-5423  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/628,417  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:









```

1  TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
2  NUMBER OF SEQUENCES:  125
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE:  CUSHMAN DABY & CUSHMAN, L.L.P.
5  STREET:  1100 New York Avenue, N.W.
6  CITY:  Washington
7  STATE:  D. C.
8  COUNTRY:  U.S.A.
9  ZIP:  20005-3918
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE:  Floppy disk
12 COMPUTER:  IBM PC compatible
13 OPERATING SYSTEM:  PC-DOS/MS-DOS
14 SOFTWARE:  PatentIn Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER:  US/08/7332,766A
17 FILING DATE:  01-NOV-1994
18 CLASSIFICATION:  435
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER:  GB 9326052.9
21 FILING DATE:  21-DEC-1993
22 ATTORNEY/AGENT INFORMATION:
23 NAME:  BIRD, Donald J.
24 REGISTRATION NUMBER:  25,323
25 REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
26 TELECOMMUNICATION INFORMATION:
27 TELEPHONE:  (202) 861-3000
28 TELEFAX:  (202) 822-0944
29 TELEX:  6714627 CUSH
30 INFORMATION FOR SEQ ID NO:  22:
31 SEQUENCE CHARACTERISTICS:
32 LENGTH:  494 base pairs
33 TYPE:  nucleic acid
34 STRANDEDNESS:  single
35 TOPOLOGY:  linear
36 MOLECULE TYPE:  DNA (genomic)
37 US-08-332-766A-22

```

[illegible]

RESULT 12  
 US-08-623-906A-18  
 : Sequence 18, Application US/08623906A  
 : Patent No. 5874217  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Stevenson, Tamara  
 :  
 : APPLICANT: Dvorak, Jan  
 :  
 : APPLICANT: Halverson, Joy  
 :  
 : TITLE OF INVENTION: Microsatellite Sequences for Canine



RESULT 14  
 PCT-US93-07261-10  
 Sequence 10, Application PC/TUS9307261  
 GENERAL INFORMATION:  
 TITLE OF INVENTION: PfEMP3 MALARIA ANTIGEN, ANALOGS, ANTIBODIES AND USES THEREOF  
 NUMBER OF SEQUENCES: 23  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: John H. C. Blasdale  
 STREET: One Giralda Farms  
 CITY: Madison  
 STATE: New Jersey  
 COUNTRY: USA  
 ZIP: 07940-1000  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: Apple Macintosh  
 OPERATING SYSTEM: Macintosh 6.0.5  
 SOFTWARE: Microsoft Word 5.1a  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US93/07261  
 FILING DATE: 19930805  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/927,531  
 FILING DATE: 07-AUG-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Blasdale, John H. C.  
 REGISTRATION NUMBER: 31,895  
 REFERENCE/DOCKET NUMBER: DX0288K  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 201-822-7398  
 TELEFAX: 201-822-7039  
 INFORMATION FOR SEQ ID NO: 10:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4766 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA to mRNA  
 ORIGINAL SOURCE:  
 ORGANISM: Plasmodium falciparum  
 STRAIN: Malayan Camp  
 IMMEDIATE SOURCE:  
 CLONE: p2b1;p12-1  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 3..4766  
 CDT-US93-07261-10

| Query Match           | 6.3%  | Score 82.4        | DB 6     | Length 4766 |
|-----------------------|---|-------------------|----------|-------------|
| Best Local Similarity | 52.0%   | Pred. NO. 3.4e-09 |          |             |
| Matches 229           | Conservative 0  | Mismatches 208    | Indels 3 | Gaps 2      |
| QY 857                | aagaaagaggaagaagagagagagaagaagaagaacaagaggaagagaaagaaga     |                   |          | 916         |
| DB 8                  | AAGAAAGCGCTTTAAACAGAAACTGCAAAAAATGAAAAAGCAGAAATGCTTAA       |                   |          | 67          |
| QY 917                | agggagagagaaagaaagaaagagagagaaagaaagagagaaagaaagagag        |                   |          | 975         |
| DB 68                 | GGAAGAAAGATTAAAGAACAAAAAAGATGCGCGAGAAAGCAAAAGATTACAA        |                   |          | 127         |
| QY 976                | aaggaagaagaaagaagagagaagaagaagaactaagaagaagagagagaaaga      |                   |          | 1035        |
| DB 128                | AAAGGAAGCCACAGACTCTTCAGTGAAAAATCATTAAAGAAAAAGTAAAGGAGAC     |                   |          | 187         |
| QY 1036               | aggaagagagaaagaaagaaagttaagaagaagaagaagaagaagaagaagagga     |                   |          | 1095        |
| DB 188                | ATTAAAGCAAAAGAAATTAAGAAACATTAAAAAAAAGATTAGAAAAACAGAAAG      |                   |          | 247         |
| QY 1096               | gaactnagaagaagaagagagagagaagaagaagaagaatataagaaacnagaagaaga |                   |          | 1155        |
| DB 248                | AAAACAGAAAAATTAATTAAGATATATATATGAAAGCATTTAA--AGATTAAGAG     |                   |          | 305         |

Qy 1156 gaagaaagatlaagaagaggaagaaagaaagaaagaaagaaagaaagaa 1215  
Db 306 AATGATTAAGATGATTAATAAATAATGTTCCCAAAAAACCGAATTCGTAAGAAAGATTG 365  
Qy 1216 aaggagaataaaaggaagaagaatagaagcggaagaagaagaagaagtataagag 1275  
Db 366 AAGAGAAATGCAATTGGAAGAGAAAGAAATTTATTAACAACATTTTAAGCGATTATGAGAG 425  
Qy 1276 gaagaagaagaagaagaa 1295  
Db 426 AGAAGGAAAAAGACAGAA 445

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

|    | Query Match           | 6.3%  | Score 82.2      | DB 5     | Length 1859 |
|----|-----------------------|---|-----------------|----------|-------------|
|    | Best Local Similarity | 52.7%   | Pred. NO. 3e-09 |          |             |
|    | Matches 174           | Conservative 0  | Mismatches 156  | Indels 0 | Gaps 0      |
| QY | 869                   | aagaagagagagagaaagaagaagaagaagaagaagaagaagaag               | 928             |          |             |
|    |                       |   |                 |          |             |
| DB | 184                   | AAAAGGGAAGAGGAGAGAGACACACTAAAGTGGCTACAGAGCGAAAGGAAACCTAGCG  | 243             |          |             |
| QY | 929                   | aagaagaagaagagagaagaagaagaagaagaagaagaagaagaaga             | 988             |          |             |
|    |                       |   |                 |          |             |
| DB | 244                   | AAAGGAAGAGAGGAAGAGACAGAAAGTCTAGAGAGAGAGAGGAAAGACACAGACACACA | 303             |          |             |



**THIS PAGE BLANK (USPTO)**





121 ttccgcaataaggtgatttcttgaatgaaatgacagagaagactatctatgat 180  
121 Ttccgcaataaggtgatttcttgaatgaaatgacagagaagactatctatgat 180  
121 TTCCGCAATAAGGTGATTCTTGAATGAAATGATGACAGAGACTATCTATGAT 180  
181 gtgcgtggaatgtacaccagaccatgtgacgtgctgtctgtggagacctgaagctg 240  
181 Gtgcgtggaatgtacaccagaccatgtgacgtgctgtctgtggagacctgaagctg 240  
181 GTGCGTGGAATGTACACCAGACCATGTGACGTGCTGTGCGGAGACTGGAAGCTG 240  
241 gtacataatgaacccagccgtctgtctgttgaatgacattcgccgtgtaccactg 300  
241 Gtacataatgaacccagccgtctgtctgttgaatgacattcgccgtgtaccactg 300  
241 GTACATAATGAACCCAGCCGTCTGTCTGTGATGACATTGCGCGCTGATGCCACTG 300  
301 aagcaacaggtggaatataatgaatcagcccccgccgtcccaaggaagctgaagagtg 360  
301 AAGCAACAGGTGGAATATGATGATGACGTGACCCCCGCGCTCCAGGAAGCTGAAGAGTG 360  
301 AAGCAACAGGTGGAATATGATGATGACGTGACCCCCGCGCTCCAGGAAGCTGAAGAGTG 360  
361 cgtcgtgacccgtctgacaccccgaaagccctcgccgtgagtgtgctgtgtggccttgagt 420  
361 Cgtcgtgacccgtctgacaccccgaaagccctcgccgtgagtgtgctgtgtggccttgagt 420  
361 Cgtcgtgacccgtctgacaccccgaaagccctcgccgtgagtgtgctgtgtggccttgagt 420  
421 gactgtgagctctcatctcccaactcaaaagccgtcgaagcagacagcgtcgagctc 480  
421 Gactgtgagctctcatctcccaactcaaaagccgtcgaagcagacagcgtcgagctc 480  
421 GACTGTGAGCTCTCATCTCCCAACTCAAAAGCCGTCAAGGCGGTCAAGACAGCTCGGGCTC 480  
481 cagtaaggagagcagatcgtccggaatcaatgtatatccattcctcctgtacccatgag 540  
481 CAGTAAGGAGAGCAGATCGTCCGGAATCAATGTATATCCATCTCCTCGTATCCCATGAG 540  
481 CAGTAAGGAGAGCAGATCGTCCGGAATCAATGTATATCCATCTCCTCGTATCCCATGAG 540  
541 gaggtcacaacctattccgaacccaagaactgtgtccatcaaaagtgaagaacctggc 600  
541 GAGGTCAACACTATTTCGAACCAAGAAACTGTGTCCATCAAAAGTGAGAACATCGGC 600  
541 GAGGTCAACACTATTTCGAACCAAGAAACTGTGTCCATCAAAAGTGAGAACATCGGC 600  
601 ctgacccccgtgaagaactctctcgaatgagcccccaacttggcaatgtatgtatgtt 660  
601 CTGACCCCCGTGAAGAACTCTCTCGAATGAGCCCCCAACTTGGCAATGTATGTATGT 660  
601 CTGACCCCCGTGAAGAACTCTCTCGAATGAGCCCCCAACTTGGCAATGTATGTATGT 660  
661 gtlctggaactctggggcgctgcgaagcagccctgggctcccttggaatctgggaaacaag 720  
661 Gtlctggaactctggggcgctgcgaagcagccctgggctcccttggaatctgggaaacaag 720  
661 GTLCTGGAACCTCTGGGCGCTGCGAAGCAGCCCTGGGCTCCCTGGAATCTGGGAAACAAG 720  
721 gagaagaaggtctcatatagcctgtgtgaagctccgaagccttggctgtgaacttccagc 780  
721 GAGAAGAAGGTCTCATATAGCCTGTGTGAGGCTCCGAGGCTTGGCTGACATTTCCAGC 780  
721 GAGAAGAAGGTCTCATATAGCCTGTGTGAGGCTCCGAGGCTTGGCTGACATTTCCAGC 780  
781 gcccccatccsaagacccctgacatcttcatcagccatgtgaaacccggtccctgtctgct 840  
781 GCCCCCATCCSAAGACCCCTGACATCTTTATCAGCCATGTGAACCTGCTCCTGTCTGCT 840  
781 GCCCCCATCCSAAGACCCCTGACATCTTTATCAGCCATGTGAACCTGCTCCTGTCTGCT 840  
841 gaggtgggaattggagataaggggacagatgtlccaagtcaatgtgcgtcgacttcttaac 900  
841 GAGGTGGGAATTGGAGATAGGGGACAGATTGCGAAGTCAATGAGCGCTTCTTAAC 900  
841 GAGGTGGGAATTGGAGATAGGGGACAGATTGCGAAGTCAATGAGCGCTTCTTAAC 900  
901 ctgataccaagggagcgtgtataatgtgtctgaaataatagccgaagcctgaaccttccatt 960  
901 Ctgataccaagggagcgtgtataatgtgtctgaaataatagccgaagcctgaaccttccatt 960  
901 CTGATACCAAGGAGCGCTGTAAATGTCTGAATAATAGCCGACGCTGACCTATCTCAT 960  
961 gtaagctgagcctgacggcgagcctgtctcatcagacagccggagccgctgtgcagaaagccgg 1020  
961 Gtaagctgagcctgacggcgagcctgtctcatcagacagccggagccgctgtgcagaaagccgg 1020  
961 GTAGCTGAGCCTGACGGCGGAGCTGTTCATGACAGACCGGAGCCGCTGTGCAGAAAGCCGG 1020  
1021 cagcgtgagcctgacggcgagcctgtctcatcagaaagccgctgtgcagatgtgagctcaac 1080  
1021 Cagcgtgagcctgacggcgagcctgtctcatcagaaagccgctgtgcagatgtgagctcaac 1080  
1021 CAGCGTGAGCCTGACGGCGGACGCTTCTCATGCGAAGCGGCTGGCATGTGAGTCCAAC 1080  
1081 aagatccctccagggagcaacagagatgtgagcggcaaaagagaagaatgtgccagaag 1140  
1081 AAGATCCCTCCAGGGAGCAACAGAGATGTGAGCGGCAAAAGAGAAGAATGTGCCAGAG 1140  
1081 AAGATCCCTCCAGGAGCAACAGAGATGTGAGCGGCAAAAGAGAAGAATGTGCCAGAG 1140  
1141 gcaagaagagaaatgaagagataccgaagagatgtgaacagatgttagagagaaagag 1200  
1141 GCAAGAAGAGAAATGAAGAGATACCGGAAGAGATGTGAACAGATGTGAGAGAGAAAGAG 1200  
1141 GCAGCAGAGGAAATGAAGAGATACCGGAAGAGATGTGAACAGATGTGAGAGAGAAAGAG 1200  
1201 aagtttaagaagcaatgaggaaagagactggggctcaaaagaaacagctactcttgccctaa 1260

1201 AAGTTAAGAAACAATGGGAAGAAAGACTGGGGCTCAAAAGAACACTACTTGGCTAAA 1260  
1261 accatcactgctgaagtagtaacccaatgaccccttgcaagcacaatgtatcttcggaata 1320  
1261 ACCATCAGTGGCTGAGAGTACACCAAGTACCCCTTGCAGACCAAAAGTATTCGGAAATA 1320  
1321 tgaaggaagccttgacccctactcatatgttcaacccagagcagatcatgaggagatgt 1380  
1321 TGAGGAAGCCTTGAACCCCTACTATGTATGTCACCCAGAGAGATCATGCGGAAGAGAT 1380  
1321 TGAGGAAGCCTTGAACCCCTACTATGTATGTCACCCAGAGAGATCATGCGGAAGAGAT 1380  
1381 ccggtcctcaagcatcaagaagaggagatccttaagacctgacctggaagcgtgtgtga 1440  
1381 CCGGCTCCTACCATCAAGAAAGAGAGGATCCTTAAGACCTGACCTTGGAAGCGGTGTGA 1440  
1381 CCGGCTCCTACCATCAAGAAAGAGAGGATCCTTAAGACCTGACCTTGGAAGCGGTGTGA 1440  
1441 ctcccccttgggaaggggtgtcttctgctgtgtgtatgagcgggagcgtgtgaacgca 1500  
1441 CTCCCCCATGGGAAGGGTGTCTGTGTGTGTATGAGCGGAGACTGTGAGCGGCA 1500  
1441 CTCCCCCATGGGAAGGGTGTCTGTGTGTGTATGAGCGGAGACTGTGAGCGGCA 1500  
1501 tegtgcattgtgaagaggagacagatcatgycacaatcaacgycacaagatgtgacagacta 1560  
1501 TEGTGCAATTGTGAAGGAGAGAGAGATCATGSCATCAACGGCAAGATTGTACAGACTA 1560  
1501 TEGTGCAATTGTGAAGGAGAGAGAGATCATGSCATCAACGGCAAGATTGTACAGACTA 1560  
1561 caacctggctgagcctgagcgtgacctgcagaaagccctggaatcagggcgggagctgtgat 1620  
1561 CAACCTGGCTGAGCCTGAGCCTGACGCTGCGCTGCAAGGCTGGAATCAAGGCGGGAGCTGAT 1620  
1561 CAACCTGGCTGAGCCTGAGCCTGACGCTGCGCTGCAAGGCTGGAATCAAGGCGGGAGCTGAT 1620  
1621 cgaactgt 1680  
1621 CGAAGT 1680  
1621 CGAAGT 1680  
1621 CGAAGT 1680  
1681 gaagtcacaagaagggaacccaataatcaacgcttgaagaaacaatgagctccgccccact 1740  
1681 GAAGTCCAAAAGGGGAACCAAAATTCACGCGTTAGGAACAATGAGCTGCCGCCACT 1740  
1681 GAAGTCCAAAAGGGGAACCAAAATTCACGCGTTAGGAACAATGAGCTGCCGCCACT 1740  
1741 cgtgaacaacaagcctcggaacagccttgaagaggagccaatgacacacacagatgagca 1800  
1741 CGTGAACAACAAGCCTCGGACAGCCTTGAGAGAGCCACATGACACACACGATGGCA 1800  
1741 CGTGAACAACAAGCCTCGGACAGCCTTGAGAGAGCCACATGACACACACGATGGCA 1800  
1801 tccctgggaacccgaatctataacccaggaatctcaaatcccttggccctgaaacagag 1860  
1801 TCCCTGGGAACCCGAATCTATACCCAGGATCTCAAACTCCCTTGGCCGTGAACCAAGG 1860  
1801 TCCCTGGGAACCCGAATCTATACCCAGGATCTCAAACTCCCTTGGCCGTGAACCAAGG 1860  
1861 ccaagaataaggaacagcctcggaacagccttltgaaggccaatgtgaggaaagagagcagc 1920  
1861 CCAAGATAAGGAACAGCTCGGGCACTTTTGAAGCCAAATGGAAGGAAGGAGCAAGC 1920  
1861 CCAAGATAAGGAACAGCTCGGGCACTTTTGAAGCCAAATGGAAGGAAGGAGCAAGC 1920  
1921 cagccgttgggagaagatctcaaggaatccagactcattcccttctgtgcccagt 1980  
1921 CAGCGTGTGGGAGAAGATCTCAAGGATCCAGACTCTATTCTTCCCTGCGCCAGTG 1980  
1921 CAGCGTGTGGGAGAAGATCTCAAGGATCCAGACTCTATTCTTCCCTGCGCCAGTG 1980  
1981 aattgtctctcccaagatttggggagctccttcttgaacccctaaagacccacgtg 2040  
1981 AATTGTCTCTCCCAAGATTGGGGAGCTCTTCTTGAACCTTAATTAAGACCCACAGTG 2040  
1981 AATTGTCTCTCCCAAGATTGGGGAGCTCTTCTTGAACCTTAATTAAGACCCACAGTG 2040  
2041 agtctctctctccatccctctctctgcccctgcgtcctaattgtctgcaggaatgtca 2100  
2041 AGTCTCTCTCTCCATCCCTCTCTCTCTGCGCTGTGCTATGTCTGCGAGGATGTGCA 2100  
2041 AGTCTCTCTCTCCATCCCTCTCTCTCTGCGCTGTGCTATGTCTGCGAGGATGTGCA 2100  
2101 ctcaaaccttactctgagctcatatataataaacaagattatcttccagcttaaaa 2160  
2101 CTCCAAACTTACTCTGAGCTCATATTAATAATAACAAGATTATTTCCAGCTTAAAAA 2160  
2101 CTCCAAACTTACTCTGAGCTCATATTAATAATAACAAGATTATTTCCAGCTTAAAAA 2160  
2161 aa 2162  
2161 AA 2162

RESULT 2  
X39644  
ID X39644 standard; DNA: 2236 BP.  
AC X39644;  
DT 02-Jul-1999 (first entry)





```
Db 661 GTGTGGAAATCTGGGGGCTGCGAGGCGACCTGGGCTCCCTGGAAATCGGGAACAAG 720
QY 721 gagaagaagttcttcataagcctgtagagctcccgaggtcttgctgcaacatltccagc 780
Db 721 GAGAAGAAGGCTTTCATCACACCTGTGTAGGCTCCCGAGGCTTGGCTGACACATTTCAGC 780
QY 781 ggcgcccatccagaagccttgcatcttatacagcccatgtgaacctgtctccctgtcgtc 840
Db 781 GGCCCCATCCAGAAAGCTGGCATCTTTATACGCCATGTGAACCTGGCTCCCTGCTGCT 840
QY 841 gaagtggaattggaagaaaggaagccagatgtgcgaatgcaatgtgcgtcgaatctctaac 900
Db 841 GAAGTGGAGATTGGAGATAGGAGGACACAGATTGTGCAAGTCAATGGCGTCGACTTCTTAAC 900
QY 901 ctggaatacaagaagagctgtcaaatgtgctgaataatagccgagcctgacatctccatt 960
Db 901 CTGGAATCAAGAGGCTGTGAAATGTGCTGAAAAATAGCCGACCCGACATCTTCATT 960
QY 961 gtgagtcgaactggtccgggagctgttcaatgaagaagccggagcggtggaagggcgcg 1020
Db 961 GTAGCTCAGACTGGCCGGGAGCTGTTTATACACACCGGAGCGGCTGGCAGAGGCGGG 1020
QY 1021 caagcgtgagctgcagcggcgagagctctcaatgcagaagcggtgcgcatggaatccaac 1080
Db 1021 CAGCGTAGCTGCGAGCGGCGAGAGCTTCTCATGCGAAGCGGCTGGCGATGGAGTCCAAC 1080
QY 1081 aagatctccagaagcagcagagagatgagagcgcaagaagaagaatltgccagaag 1140
Db 1081 AAGATCTCCAGAGCGAGCGAGAGATGAGACGGCAAGGAAAGAAATTTGCCCAACAG 1140
QY 1141 ggcagcaagaagaatgagaatatacccggaagagatggaacagatltgagaggaaggaag 1200
Db 1141 GCAGCAGAGGAAATGAGAAATACCGGAAAGAGATGGAACAGATTGTAGAGGAAGAGAG 1200
QY 1201 aaglttaagaagaatggaagaagacttgagggtcaagaagaacagctactccttgctaa 1260
Db 1201 AAGTTTAAGAAGCAATGGGAAGAGACTGGGGCTCAAGAGAACAGACTCTTGCCATAA 1260
QY 1261 accataactgtcgtaggtatacaccagtaacccttcgcgaagccaaagt----- 1307
Db 1261 ACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCAAAAGTATGATCAGGAGTG 1320
QY 1307 ----- 1307
Db 1321 GAACCTGAGCTCGAGCCCGGAGATGACCTGAGTAGAGGACGACGAGAGAGACAG 1380
QY 1307 gattccggaatatagagaagagcttgaccctactctatgttcaaccacagacagatc 1366
Db 1381 GATTTCGGAAATATGAGGAAGGCTTGACCCGACTCTATGTTTCAACCCAGAGCAGATC 1440
QY 1367 atgagggaaggaatgctccggtcctcaagcatcaagaagaagagatccttaagacctggc 1426
Db 1441 ATGGGGGAAGGATGCCGGCTCTCTACGCATCAGAGAGAGGAGATCTTATGACGCGCCTG 1500
QY 1427 gaagggggtgtgagacctcccatatggaagaagtgtgttctgtcgtgtagtgcgggga 1486
Db 1501 GAAGGGGgtgtgagacctcccatatggaagaagtgtgttctgtcgtgtagtgcgggga 1560
QY 1487 gctgctgaagcagatggtgcatgttgaagaaggaagacagatcaatgaaatcaacggaag 1546
Db 1561 GCTGCTGAAGGCGCATGTTGCGATTGTCAAAAGGGGACGAGATCATGCGAATCAACGGCAG 1620
QY 1547 atttggaacagactaacacctgtgtaggtgagcgtgccttgcaagaagcgctggaatcag 1606
Db 1621 ATTGTGACAACTAACACCTCGCTGAGGCTGACGCTCCCTGCAGAAAGCGCTGGAATCAG 1680
QY 1607 gggggggagcagatgagacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1666
Db 1681 GGGGGGgagcagatgagacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740
QY 1667 ctgacctctctgtgaagtcacaaaggggaacaaatcaacagcgttaagaaacagttag 1726
Db 1741 CTGACCTTCTTGTGTAAGTCCAAAGGGGAACCAATTATCACCGTTAGGAACAGTGAG 1800

QY 1727 ctccggcccaactcgtgaaacaaaagcctcggaccagccttgagagagggccaatgaca 1786
Db 1801 CTCGGGCCCCACCTCGTGAACACAAAGCTTCGGACCGACTTGAAGAGAGGCGACATGACA 1860
QY 1787 cacacagatggaatctcttggaacctgaatctataccaggaatcccaactcccttg 1846
Db 1861 CACACGAGATGGCATCTTGGGACCTGATCTTACACCCAGAAATCTCAAACTCCCTTGG 1920
QY 1847 gccctgaaccaaagggccagataaaggaacagctcggccacttlttgaaggccaatgga 1906
Db 1921 GCCCTGAACCAAGGGCCAGATTAAGGAACAGCTCGGGCCACTTTTGTGAAGCCAAATGGA 1980
QY 1907 ggaagggagcagcagcgcttggggaagaatctcaagaatccagaatctcaatctt 1966
Db 1981 GGAAGGGAGCACACCACTTGGGGAAGATCTCAAGGATCCAGACTCTTCATTCTCTT 2040
QY 1967 cctctggccagtgaaattgtctctccagccttgggagactccctccttgaaacctaa 2026
Db 2041 CCTCTGGCCAGTGAATTTGTGTCTCTCCAGCTTTGGGGGACTCTTCTTGAAACCTTA 2100
QY 2027 taagaaccacatggaatctctctctcaatccctcctccttgcccttgcttaattgc 2086
Db 2101 TAAGACCCCACTGGAGTCTCTCTCTCCATCCCTCTCTGCGCTCTGCTCTTAATTGC 2160
QY 2087 tgcaggaattgtactctcaaaccttactctgagctcaataataaacaagattat 2146
Db 2161 TGCAGGATTTGTCACTCCAAACCTTACTGTGAGCTCATTAATAAATAACAGATTATT 2220
QY 2147 ttccagcttaaaaaa 2162
Db 2221 TTCCAGCTTAAAAAAA 2236

RESULT 4
X40061
ID X40061 standard: DNA; 2289 BP.
AC X40061;
DT 02-JUL-1999 (first entry)
DE Colon cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; SS.
OS Homo sapiens.
PN W09J04265-A2.
PD 28-JAN-1999.
PE 15-JUL-1998; U14679.
PR 22-JUN-1998; US-102322.
PR 17-JUL-1997; US-896164.
PR 10-OCT-1997; US-061599.
PR 10-OCT-1997; US-061765.
PR 10-OCT-1997; US-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Cure A, Ohare M, Ohta Y, Old LJ,
PI Freundschuh M, Sahin U, Scanlan MJ, Stockert E,
PI Tureci O;
DR WPI; 99-13248/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Claim 67; Page 663-664; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
```

CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.

Sequence 2289 BP; 567 A; 596 C; 668 G; 458 T;

Query Match 93.7%; Score 2025; DB 1; Length 2289;

Best Local Similarity 94.5%; Pred. No. 0;

Matches 2162; Conservative 0; Mismatches 0; Indels 127; Gaps 1;

```
OY 1 cctggcccgctgctgctgctgcttcccaactccttgcaacccgggaccccggaagac 60
Db 1 cctggcccgctgctgctgctgcttcccaactccttgcaacccgggaccccggaagac 60
OY 61 gggctgctgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 120
Db 61 gggctgctgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 120
OY 121 ttcggcaataaggctgctgctgctgctgctgctgctgctgctgctgctgctgct 180
Db 121 ttcggcaataaggctgctgctgctgctgctgctgctgctgctgctgctgctgct 180
OY 181 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 240
Db 181 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 240
OY 241 gtcacataaagacacacacacacacacacacacacacacacacacacacacac 300
Db 241 gtcacataaagacacacacacacacacacacacacacacacacacacacacac 300
OY 301 aagcaacagctggaataatgatacagctgaccccccgcgcgtccagaagctgaaggagtg 360
Db 301 aagcaacagctggaataatgatacagctgaccccccgcgcgtccagaagctgaaggagtg 360
OY 361 cgtcggacacgtctgcaaccccggaagcctcgccgctgctgctgctgctgctgctgct 420
Db 361 cgtcggacacgtctgcaaccccggaagcctcgccgctgctgctgctgctgctgctgct 420
OY 421 ggcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 480
Db 421 ggcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 480
OY 481 caggtgaaggagacagacagctgctgctgctgctgctgctgctgctgctgctgctgct 540
Db 481 caggtgaaggagacagacagctgctgctgctgctgctgctgctgctgctgctgctgct 540
OY 541 gaggtacacacacacacacacacacacacacacacacacacacacacacacacac 600
Db 541 gaggtacacacacacacacacacacacacacacacacacacacacacacacacac 600
OY 601 ctgaccccgctggaagcctcctcctgctgctgctgctgctgctgctgctgctgctgct 660
Db 601 ctgaccccgctggaagcctcctcctgctgctgctgctgctgctgctgctgctgctgct 660
OY 661 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 720
Db 661 gctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 720
OY 721 gagaagaagcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 780
Db 721 gagaagaagcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 780
OY 781 ggcacacacacacacacacacacacacacacacacacacacacacacacacacac 840
Db 781 ggcacacacacacacacacacacacacacacacacacacacacacacacacacac 840
OY 841 gaggtggaatggaagataagggaacagatgctgcaagtaacatgagctgacacttctaac 900
Db 841 gaggtggaatggaagataagggaacagatgctgcaagtaacatgagctgacacttctaac 900
OY 901 ctggtacacagagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 960
Db 901 ctggtacacagagagctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 960
```

```
OY 961 gtagctgcaagctgcccggagagctgttcatagacacacccggagcggctgcaagagcgcg 1020
Db 961 gtagctgcaagctgcccggagagctgttcatagacacacccggagcggctgcaagagcgcg 1020
OY 1021 cagcgtgagctgcaagcaggaagagcttctcctcctcctcctcctcctcctcctcct 1080
Db 1021 cagcgtgagctgcaagcaggaagagcttctcctcctcctcctcctcctcctcctcct 1080
OY 1081 aagatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1140
Db 1081 aagatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1140
OY 1141 gcaagcaggaataatgagataacccggaagagatggaacacattgtagaggaagagag 1200
Db 1141 gcaagcaggaataatgagataacccggaagagatggaacacattgtagaggaagagag 1200
OY 1201 aagtttaagaagcaatggaagaagactgaggctcacaagaagacactccttgctaa 1260
Db 1201 aagtttaagaagcaatggaagaagactgaggctcacaagaagacactccttgctaa 1260
OY 1261 accatcactgctggaagtaacacacacacacacacacacacacacacacacacacac 1307
Db 1261 accatcactgctggaagtaacacacacacacacacacacacacacacacacacacac 1307
OY 1307 -----gattcc 1313
Db 1307 -----gattcc 1313
OY 1381 ccacagagagatgtaagagagatggtggtttatcaagacacagattcaagacacagattcc 1440
Db 1381 ccacagagagatgtaagagagatggtggtttatcaagacacagattcaagacacagattcc 1440
OY 1314 ggaataatgagaagagcttgaacccctactctatcttaaccccaagaagatcagagga 1373
Db 1314 ggaataatgagaagagcttgaacccctactctatcttaaccccaagaagatcagagga 1373
OY 1441 ggaataatgagaagagcttgaacccctactctatcttaaccccaagaagatcagagga 1500
Db 1441 ggaataatgagaagagcttgaacccctactctatcttaaccccaagaagatcagagga 1500
OY 1374 agagctgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1433
Db 1374 agagctgctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1433
OY 1501 aggatgctcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1560
Db 1501 aggatgctcctgctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1560
OY 1434 gctgagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1493
Db 1434 gctgagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1493
OY 1561 gctgagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1620
Db 1561 gctgagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1620
OY 1494 agcgcacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1553
Db 1494 agcgcacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1553
OY 1621 agcgcacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1680
Db 1621 agcgcacatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1680
OY 1554 cagactacacccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1613
Db 1554 cagactacacccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1613
OY 1681 cagactacacccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1740
Db 1681 cagactacacccctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1740
OY 1614 actgatacagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1673
Db 1614 actgatacagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1673
OY 1741 actgatacagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1800
Db 1741 actgatacagcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1800
OY 1674 tcttgctgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1733
Db 1674 tcttgctgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1733
OY 1801 tcttgctgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1860
Db 1801 tcttgctgagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1860
OY 1734 cccacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1793
Db 1734 cccacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1793
OY 1861 cccacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1920
Db 1861 cccacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1920
OY 1794 gatgacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1853
Db 1794 gatgacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1853
OY 1921 gatgacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1980
Db 1921 gatgacacacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1980
OY 1854 accagagccagataagagacagctcgggacacatttttgaagggcaatgctgagaggaag 1913
Db 1854 accagagccagataagagacagctcgggacacatttttgaagggcaatgctgagaggaag 1913
OY 1981 accagagccagataagagacagctcgggacacatttttgaagggcaatgctgagaggaag 2040
Db 1981 accagagccagataagagacagctcgggacacatttttgaagggcaatgctgagaggaag 2040
```



|    |      |   |      |
|----|------|---|------|
| QY | 1914 | gagcagccagccgctttgggaggaagatctcaaggatccagactctatcttccttcctgtg | 1973 |
| Db | 2041 | GAGCAGCCAGCCCTTTGGGAGGAAGATCTCAAGGATCCAGACTCTATCTTCCTCCG      | 2100 |
| QY | 1974 | cccgatgaattgtgtctctccagcctttgggagactcctctcttgaaocccaataaagac  | 2033 |
| Db | 2101 | CCCGATGAATTTGTGTCTCTCCAGCTTTTGGGGAGCTCCTCTTAAACCTAATAAGAC     | 2160 |
| QY | 2034 | ccactggagctctctctctcaatccctctcctctgtccctgtctaatgtgtgcag       | 2093 |
| Db | 2161 | CCACTGGAGCTCTCTCTCTCAATCCCTCTCCTCTGTCTAATGTGTCGAG             | 2220 |
| QY | 2094 | atgtgcactccaaaccttactctgagctcatatataaataaacaagatattatccagc    | 2153 |
| Db | 2221 | ATGTGCACCTCAAAACCTTACTCTGAGCTCATTAATAATAACAGATTTATTTCCAGC     | 2280 |
| QY | 2154 | ttaaaaaa 2162   |      |
| Db | 2281 | TTAAAAAAA 2289  |      |

|        |   |
|--------|---|
| RESULT | 5   |
| ID     | X40062 standard; DNA; 2409 BP.  |
| AC     | X40062;   |
| DT     | 02-JUL-1999 (first entry)   |
| DE     | Colon cancer associated gene.   |
| KW     | Cancer associated antigen; diagnosis; research; treatment; human;<br>breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;<br>prostate cancer; ss.  |
| OS     | Homo sapiens.   |
| PN     | WO9904265-A2.   |
| PD     | 28-JAN-1999.  |
| PF     | 15-JUL-1998; UI4679.  |
| PR     | 22-JUN-1998; US-102322.   |
| PR     | 17-JUL-1997; US-896164.   |
| PR     | 10-OCT-1997; US-061599.   |
| PR     | 10-OCT-1997; US-061765.   |
| PR     | 10-OCT-1997; US-948705.   |
| PR     | 11-OCT-1997; GB-021697.   |
| PA     | (LUDMIG) INST CANCER RES.   |
| PI     | Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,   |
| PI     | Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,<br>Tureci O;   |
| DR     | WI: 99-132448/11.   |
| PT     | New isolated cancer associated nucleic acids and polypeptides -   |
| PT     | isolated using sera from cancer patients, used to develop products<br>for the diagnosis, monitoring or treatment of cancers   |
| PS     | Claim 67; Page 665-666; 787pp; English.   |
| CC     | The invention relates to a method for diagnosing a disorder characterised<br>by expression of a human cancer associated antigen precursor coded for by<br>a nucleic acid molecule (NAM). The method comprises: (a) contacting a<br>biological sample isolated from a subject with an agent that specifically<br>binds to the NAM, an expression product or a fragment of an expression<br>product complexed with an HLA molecule; and (b) determining the<br>interaction between the agent and the NAM or the expression product as a<br>determination of the disorder. The products and methods can be used in<br>the diagnosis, monitoring, research, or treatment of conditions<br>characterised by the expression of various cancer associated antigens.<br>The invention provides nucleic acid sequences and encoded polypeptides<br>which are cancer associated antigen precursors expressed in human breast<br>cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and<br>lung cancer. |
| Q0     | Sequence 2409 BP; 595 A; 638 C; 692 G; 484 T;   |

|                            |        |              |       |                     |
|----------------------------|--------|--------------|-------|---------------------|
| Query Match                | 88.1%; | Score 1905;  | DB 1; | Length 2409;        |
| Best Local Similarity      | 89.7%; | Pred. No. 0; |       |                     |
| Matches 2162; Conservative | 0;     | Mismatches   | 0;    | Indels 247; Gaps 1; |

**OY**      1 cctggcccggtcgcgctccttccaagtcctfycagccgycacaccgaagaac 60  
         |||||  
**D8**      1 CCTGCCCGGCTCGGGTCCGGCTTTTCACACTCCTGGCAAGCCGGACCCGAGGAAAC 60

|    |      |  |      |
|----|------|--|------|
| QY | 61   | gggtctcgtgcaacgcagcgcagctctggaactggtcccgcaatgagccgaaagtggcccgagaa      | 120  |
| Db | 61   | gggtctgtrcaacgcagcgcagctctggaactggtcccgcaatgagccgaaagtggcccgagaa       | 120  |
| QY | 121  | ttccgagcaatgagtgatatttcctgattgaaatgatgagcagaagaagcattctcatgat          | 180  |
| Db | 121  | ttccgagcaatgagtgatatttcctgattgaaatgatgagcagaagaagcattctcatgat          | 180  |
| QY | 181  | gtgctgcgaatctgaaccaaccaatgagctgagccgtctcgttgggagacctgaactg             | 240  |
| Db | 181  | gtgctgcgaatgagtgatgacacacagacctgagccgtctcgttgggagacctgaactg            | 240  |
| QY | 241  | gtcatcaatgagaccagccgtctgcccctgtttgatgacatttcggccgttgatctccaatg         | 300  |
| Db | 241  | gtcatcaatgagaccagccgtctgcccctgtttgatgacatttcggccgttgatctccaatg         | 300  |
| QY | 301  | aagcaccagctgagaaatgaatgcagctgagcccccgcgtccagaagctgaagagagtg            | 360  |
| Db | 301  | aagcaccagctgagaaatgagatgagctgagcccccgcgtccagaagctgaagagagtg            | 360  |
| QY | 361  | cgctctgagccgtctgaccccccggaagcctgagctgagtgctgctggtggtgacctgaatt         | 420  |
| Db | 361  | cgctctgagccgtctgaccccccggaagcctgagctgagtgctgctggtggtgacctgaatt         | 420  |
| QY | 421  | gagctgagcctctcatctccacacctatcaagagcggtcgaagcaagacgctgagctc             | 480  |
| Db | 421  | gagctgagcctctcatctccacacctatcaagagcggtcgaagcaagacgctgagctc             | 480  |
| QY | 481  | cagtgaggagcagatcgtctcgtgataatgatatcatctcctcctctgatacccatgag            | 540  |
| Db | 481  | cagtgaggagcagatcgtctcgtgataatgatatcatctcctcctctgatacccatgag            | 540  |
| QY | 541  | gaggtcatcaaccctcatttgtaaccaagaagaactgttcataaagtgaagaacatggcc           | 600  |
| Db | 541  | gaggtcatcaaccctcatttgtaaccaagaagaactgttcataaagtgaagaacatggcc           | 600  |
| QY | 601  | ctgataccccctgaaaaagcctctccctgataagccccctcaacttggcagatattgatacagtt      | 660  |
| Db | 601  | ctgataccccctgaaaaagcctctccctgataagccccctcaacttggcagatattgatacagtt      | 660  |
| QY | 661  | gtctcagaaatctggtggcgctgagagcgagcctgagctccctctgaaatctcgggaaaaaag        | 720  |
| Db | 661  | gtctcagaaatctggtggcgctgagagcgagcctgagctccctctgaaatctcgggaaaaaag        | 720  |
| QY | 721  | gagagaagaggtcttcatacaacgctgtagagctctcccgagagccttggctgtaagcaattccagc    | 780  |
| Db | 721  | gagagaagaggtcttcatacaacgctgtagagctctcccgagagccttggctgtaagcaattccagc    | 780  |
| QY | 781  | ggccccacacagaagcctggaatcttatacagccatgtaaaccttgcctccctgtctgct           | 840  |
| Db | 781  | ggccccacacagaagcctggaatcttatacagccatgtaaaccttgcctccctgtctgct           | 840  |
| QY | 841  | gagtgtagattgtagataaaggaggacagagattgttgaagtaaatgaggttgaacttctaac        | 900  |
| Db | 841  | gagtgtagattgtagataaaggaggacagagattgttgaagtaaatgaggttgaacttctaac        | 900  |
| QY | 901  | ctgagatcaacaagagagctgtgaatatgtgctgaataaataagccgagccttgaaccaatctccatt   | 960  |
| Db | 901  | ctgagatcaacaagagagagctgtgaatatgtgctgaataaataagccgagccttgaaccaatctccatt | 960  |
| QY | 961  | gttagctgagctgagccgggagctgcttccatgaacagaccgggagagcgctgagagagcgcg        | 1020 |
| Db | 961  | gttagctgagctgagccgggagctgcttccatgaacagaccgggagagcgctgagagagcgcg        | 1020 |
| QY | 1021 | cagcgtgagcctgagccggagagagcttctcatgacagaacggcctgagcgtgagctgagcgaac      | 1080 |
| Db | 1021 | cagcgtgagcctgagccggagagagcttctcatgacagaacggcctgagcgtgagcgtgagcgaac     | 1080 |
| QY | 1081 | aagatccctcaagagcagcagagatgtagcggccaagagagaanaaatgtgcccaag              | 1140 |
| Db | 1081 | aagatccctcaagagcagcagagatgtagcggccaagagagaanaaatgtgcccaag              | 1140 |

Db 1081 AGATCCTCAGGAGCAGCAGGATGAGCGGCAAGAAGAAATTGCCAGAG 11400

```
QY 1141 gcacagaggaataatgatagatataccggaagagatggaacagatgttagaggaggaaag 1200
    |||
Db 1141 GCAGCAGAGGAATAATGATAGATATACCGGAAGAGATGGAACATATTGTAGAGAGAGAGG 1200
QY 1201 aagtttaagaagaacatgggaagaagaactgaggctcaaaagaacagctactcttgcctaa 1260
    |||
Db 1201 AAGTTTAAGAAACAAATGGGAAGAAAGACTGGGCTCAAAAGAACACTACTCTTGTCTTAA 1260
QY 1261 accatcactgtcgtgaagtagacacccagtagcccccttgcgaagccaaagt----- 1307
    |||
Db 1261 ACCATCTACTGCTGAGGTATACCCAGTACCCCTTTCGAAAGCAAGTATGATCAGGAGATG 1320
QY 1307 ----- 1307
Db 1321 GAACCTGAGCTGAGCCCGCAGATGACCTGATGAGAGGACGAGGAGAGGAGAGCAG 1380
QY 1307 ----- 1307
Db 1381 ACATTTTGCCCAAGCCCAAGCCTCCAGAGGCCCTGGCGTGTCCACATCTCAAAACCT 1440
QY 1307 ----- 1307
Db 1441 GTCATGTGCCACAGAGAGCCCAATTTCACTACAGGCCAGCTGTGAATCTGAAGTTCTG 1500
QY 1307 ----- 1307
Db 1501 CCACAGGAGATGTGAGAGATGATGGTGGTTTATCAAGACAGCATTCAGACAAAGATTTC 1560
    |||
QY 1314 ggaataatgaggaaggctttagaccctactctagtltcaecccagaagcatcatgaggaa 1373
    |||
Db 1561 GGAATATGAGAGAGAGGCTTTGACCCCTACTCTATCTACCCCAAGAGCAGATCAAGGGGGA 1620
QY 1374 aggaatgcgcgtctctacacatcaagaagaagaagatcccttagacctggccctggaggag 1433
    |||
Db 1621 AGGATGTCCGGCTCTACGCATCAAGAGAGAGGAGATCCTTAAACCTGGGCCCTGGAGAGGG 1680
QY 1434 gctgtgactcccccaactggaagagtgctcgttcttcgctgtgtatagacgagggagctgtg 1493
    |||
Db 1681 GTGTGACTCCCCCATTTGGGAAGGTGTGCTTTCGTGTGTATGAGCGGGAGAGCTGCTG 1740
QY 1494 agcgcatgtgtgacatctgtgaaagggaacgagatcatgtgcaatcaacgagcaagattgtga 1553
    |||
Db 1741 AGCGCATGTGTGCAATTTGGAAGGGGAGAGAGATCATGGCAATCAACGCAAGATTGTGA 1800
QY 1554 cagacttaacccctgtgactaggtgacgtgacctgcaagaagccgtgagatcagggcgagg 1613
    |||
Db 1801 CAGACTTAACCTGGCTGAGGCTGAGCGCTGCTCTCAGAGAGCGTGGAAATCAAGGCGGGG 1860
QY 1614 actgtagcagctgtgtgtgctgcctgcctcccccagaagaagatagcagtagctgcacct 1673
    |||
Db 1861 ACTGATGTGACCTTGTGGTGTGCGCTGCTGCCCCCAAGAGATGATGACATGAGCTGACCT 1920
QY 1674 tctgtctgaagtccaagaagggaacccaatcacgcgttaggaacagtgagctcgagc 1733
    |||
Db 1921 TCTTCTGAAGTCCAAAAGGGGGAACCAAAATTCACGCGTTAGAGAAACAGTGAAGCTCCGCG 1980
QY 1734 cccacctgtgaacaacaaagctgagcagacgtgagaagagggagagagagagagagagagag 1793
    |||
Db 1981 CCCACCTGTGTAACACAAAGCCTCGGACAGCCTTGAAGAGAGGCCACATGATACACACCA 2040
QY 1794 gatggcatcctctgagacctgaatcatcatcaccagaagaatccaactcctcttgccctga 1853
    |||
Db 2041 GATGGCATCTTGTGGAGCTGATATATCACCAGGAATCTCAAAATCTCCCTTTGGCGCTGA 2100
QY 1854 accagggccagataaagaacagctcgagccacttcttgaagggcaatgtgaggaagaa 1913
    |||
Db 2101 ACCAGGGCCAGATAAGAAACAGCTCGGGCCACTTTTGAAGGCCCAATGTGAGAGAAAGG 2160
QY 1914 gagcagcgaagccgttgggagagaagatctcaagagatccagactctcatctctctcgtg 1973
    |||
Db 2161 GAGCAGCCAGCCGTTTGGAGAAAGATCTCAAGGATCCAGACTCTCATCTTCTTCCCTCGG 2220
QY 1974 cccagtgaaattgtgtctctccagcttggggagactcctccttgtaacccctaataagacc 2033
```

```
Db 2221 CCCAGTGAAATTTGTCCTCCACAGCTTTGGGGAGCTCCTTCCCTGACCCCTATATAGACC 2280
    |||
QY 2034 ccactggagctctctctctccatccctctctcgtgcctcgtccttaattgctgcagg 2093
    |||
Db 2281 CCAGTGGAGTCTCTCTCTCCATCCCTCTGCTGCTGCTCTAATTCCTCCAGG 2340
QY 2094 atgtgcactccaacacttactctgtgcatatataaataacagatattctccagc 2153
    |||
Db 2341 ATTGTCACTCCAAACCTTACTGTGAGCTCATTTAATAAATAACAGATTATTTCCAGC 2400
QY 2154 ttaaaaaa 2162
    |||
Db 2401 TTTAAAAA 2409
RESULT 6
X40060
ID X40060 standard; DNA; 1306 BP.
AC X40060;
DE 02-JUL-1999 (first entry)
DE Colon cancer associated gene.
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
OS Homo sapiens.
PN M09904265-AZ.
PD 28-JAN-1998.
PF 15-JUL-1998; U14679.
PR 22-JUN-1998; U5-102322.
PR 17-JUL-1997; U5-896164.
PR 10-OCT-1997; U5-061599.
PR 10-OCT-1997; U5-061765.
PR 10-OCT-1997; U5-948705.
PR 11-OCT-1997; GB-021697.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ,
PI Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E,
PI Tureci O.
DR WPI: 99-132448/11.
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
PS Claim 67: Page 662; 787pp; English.
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.
SQ Sequence 1306 BP; 348 A; 331 C; 375 G; 252 T;
```

Query Match 53.1%; Score 1148; DB 1; Length 1306;

Best Local Similarity 94.3%; Pred. No. 3.2e-295;

Matches 1232; Conservative 0; Mismatches 0; Indels 74; Gaps 1;

QY 931 aaaaatagccgagcctgacatctcatctgtagctgagctggcggagagctgtcatg 990

Db 1 AAAAATAGCCGAGCGCTACCATCTCCATTGTAGCTGACGCTGGCGGAGCTGTCTATG 60

QY 991 acagacccggagcggctggtgcagagagcgcgagcgtgtgagctgacgagcgagagcttc 1050

Db 61 ACAGACCGGAGAGCGGCTGGCAGAGGCGCGCAGCTGAGCTGACGCGCAGAGAGCTTCTC 120





```
RESULT 8
V19941/c
ID V19941 standard; DNA: 137507 BP.
AC V19941:
DE 03-AUG-1998 (first entry)
KW KSHV long unique coding region and terminal repeat.
KW KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
complement-binding protein; glycoprotein; capsid protein IV; infection;
immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
lymphoproliferative disease; leukemia; splenomegaly; mycosis fungoides;
Kaposi's sarcoma-associated herpes virus.
OS HIV immune status; anti-inflammatory agent; therapy; ds.
FH Key
FT CDS
      /tag= a
      /product= complement-binding protein
      8699..11236
      /tag= b
      /product= glycoprotein B
      complement (17261..17875)
      /tag= c
      /product= interleukin 6
      complement (21548..21832)
      /tag= d
      /product= macrophage inflammatory protein II
      complement (27137..27424)
      /tag= e
      /product= interferon regulatory factor 1
      28661..29741
      /tag= f
      /product= protein TI.1
      complement (58976..60175)
      /tag= g
      /product= glycoprotein M
      complement (69412..69915)
      /tag= h
      /product= glycoprotein L
      complement (88410..88910)
      /tag= i
      /product= interferon regulatory factor 2
      89600..90541
      /tag= j
      /product= interferon regulatory factor 3
      90173..90643
      /tag= k
      /product= glycoprotein X
      complement (93636..94127)
      /tag= l
      /product= interferon regulatory factor 4
      complement (111931..112443)
      /tag= m
      /product= capsid protein IV
      complement (123808..127296)
      /tag= n
      /product= immediate early protein
      90804576-A1.
      05-FEB-1998.
      22-JUL-1997. U13346.
      29-NOV-1996. US-757669.
      25-JUL-1996. US-686243.
      25-JUL-1996. US-686349.
      25-JUL-1996. US-686350.
      25-JUL-1996. US-687253.
      25-JUL-1996. US-688814.
      05-SEP-1996. US-708678.
      10-OCT-1996. US-728323.
      13-NOV-1996. US-747887.
      13-NOV-1996. US-748640.
      (UYCO.) UNIV COLUMBIA NEW YORK.
      PA Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
      PI WPI; 98-130615/12.
      PT New nucleic acid encoding Kaposi's sarcoma associated herpes virus
```

```
PT proteins - useful for, e.g. detecting levels of HHV8 in, and
PR preparation of vaccines for treatment of, HIV patients
PS Example 2: Page 135-203; 230pp; English.
CC This sequence represents the long unique region and terminal repeat of
the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
invention which encode KSHV polypeptides selected from: (a) viral
macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
(c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
(d) capsid protein IV encoded by ORF5; and (e) immediate early protein
encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
by it, and antibodies (Ab) specific for the proteins are useful for
detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
fluids or tissue samples. HHV8 infections can be treated with antisense
or triplex forming molecules or agents that bind specifically to the
protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
while the protein can be used in protective vaccines. Ab may also be used
to differentiate between lymphomas, and HHV8 may be implicated in many
other lymphoproliferative diseases such as lymphomas, leukemia,
splenomegaly and mycosis fungoides. Cells and animals containing the
nucleic acid are useful for drug screening. HHV8-derived nucleic acids
can be used as targets for antiviral drugs. e.g. dihydrofolate reductase gene
can be inhibited with methotrexate. These can also be used to determine
the immune status of a patient infected with HIV. HHV8 derived protein
CC viral MIP III may be used as an anti-inflammatory agent for,
CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
81 open reading frames.
SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T;
```

```
Query Match 2.7%; Score 58.8; DB 1; Length 137507;
Best Local Similarity 52.4%; Pred. NO. 4.7e-05;
Matches 129; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
```

```
QY 978 ggaagctgttcataagacagcggagcgcgtgcaagaagcgcgcgtgagctgcaagc 1037
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125113 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 125054
QY 1038 ggaagagcttcacatgacagcagcgcgtgcaagaagcgcgcgtgagctgcaagc 1097
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125053 GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 124994
QY 1098 gcaagagatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 124993 GTTACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 124934
QY 1158 gagataccggaagagatgagacagatcgtgagagagagagagagagagagagagc 1217
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 124933 GGAGTTAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 124874
QY 1218 ggaaga 1223
    |||||
DB 124873 GCAGCA 124868
```

```
RESULT 9
V55831
ID V55831 standard; DNA: 799 BP.
AC V55831:
DE 18-NOV-1998 (first entry)
DT Nucleotide sequence of the stabilising sequence-encoding insert.
KW Fusion protein; stabilising polypeptide; proteolytic degradation;
resistance; half-life; autoimmune disease; inflammation; nitro drug;
IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;
nitroreductase protein; enzyme therapy; prodrug therapy; protease;
KW cancer; pathological condition; ss.
OS Epstein-Barr virus.
FN WO9822577-A1.
PD 28-MAY-1998.
PR 17-NOV-1997; IB1508.
PR 25-JUN-1997; US-048945.
PR 15-NOV-1996; US-030986.
PA (MASU/) MASUCI M G.
```

PI Masucci MG;  
 DR WPI: 98-312463/27.  
 PT New fusion proteins resistant to proteolytic degradation -  
 PT a peptide sequence containing glycine repeats  
 PS Disclosure: Fig 4b; 120pp; English.  
 CC This is a nucleotide sequence of the stabilizing sequence-encoding  
 CC insert. The invention provides a method for increasing the resistance  
 CC of a core protein to proteolytic degradation that comprises linking or  
 CC inserting onto or into the core protein a stabilizing polypeptide of  
 CC formula (Gly)X(Gly)Y(Gly)Z/n where Glya, Glyb, Glyc are 1-6  
 CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,  
 CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not  
 CC be identical from n repeat to n repeat. Alternatively a nucleic acid  
 CC encoding the stabilizing polypeptide can be linked onto or inserted into  
 CC a nucleic acid encoding a core protein. The fusion proteins of the  
 CC invention are more resistant to degradation by proteases and, thus, have  
 CC a longer half-life than the unfused core protein. The products can be  
 CC used for treating autoimmune diseases, cancer and inflammation. In  
 CC particular, the core protein may be an IkappaB regulator protein for the  
 CC treatment of inflammatory bowel disease, or a nitroreductase protein  
 CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  
 CC or other pathological conditions. The fusion proteins can also be used in  
 CC diagnostic methods such as in vivo imaging.  
 SQ Sequence 799 BP; 201 A; 106 C; 479 G; 13 T;

Query Match 2.6%; Score 57.2; DB 1; Length 799;  
 Best Local Similarity 50.7%; Pred. No. 9.5e-06;  
 Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

OY 976 cggagactgttcatacagaccggagcgctgacagagcgcgacgctgagctgcag 1035  
 Db 42 CGGAGAGGCGAGGAGGAGGCGGCGAGGAGGAGGCGGAGGAGGAGGAGGCGG 101  
 OY 1036 cggagagacttctcattcagaagcgctgagatgagtcacaaagatcccccagag 1095  
 Db 102 CAGGAGGCGCAGAGAGGCGGCGAGGCGGCGAGGAGGCGGCGAGGAGGAGG 101  
 OY 1096 cagcagagagatggagcgcaagaagaaatggtccagaagcgacgagagaat 1155  
 Db 162 GGGCAGAGGCGGCGAGGCGGCGAGGAGGAGGCGGCGAGGAGGAGGCGG 221  
 OY 1156 gagagatcccggaagagatggaacagattgtgagagagagaagtttaagaagcaa 1215  
 Db 222 GAGGGGCGAGGAGGAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGG 281  
 OY 1216 tgggaagaagactggggtcctcaagaagacag 1245  
 Db 282 CAGGAGCAGAGGAGGCGGCGAGGAGGCGGCGAG 311

RESULT 10  
 V21683  
 ID V21683; standard; DNA; 9600 BP.  
 AC 17-AUG-1998 (first entry)  
 DE Vector plasmid pCMVkmITR-EPI.  
 KW Polynucleotide delivery; plasmid pCMVkmITR-EPI; vector;  
 KW gene therapy; vaccine; polycationic agent; ss.  
 OS Chimeric - Epstein-Barr virus.  
 OS Chimeric - Adeno-associated virus.  
 OS Chimeric - Cytomegalovirus.  
 OS Chimeric - Bos taurus.  
 FH Location/Qualifiers  
 FT Key 14..2594  
 FT CDS  
 FT /tag= a  
 FT /product= "EBV nuclear antigen A"  
 FT misc\_feature 2623..4559  
 FT /tag= b  
 FT /note= "EBV origin of replication"  
 FT repeat\_unit 4928..5104  
 FT /tag= c

FT /rpt\_type= INVERTED  
 FT /note= "AAV inverted terminal repeat"  
 FT repeat\_unit 7189..7355  
 FT /tag= d  
 FT /rpt\_type= INVERTED  
 FT /note= "AAV inverted terminal repeat"  
 FT promoter 5112..6734  
 FT /tag= e  
 FT /note= "CMV immediate-early enhancer/promoter"  
 FT terminator 6818..7050  
 FT /tag= f  
 FT /note= "bovine growth hormone polyA sequence"  
 PD WO9806437-A2.  
 PD 13-FEB-1998.  
 PF 13-AUG-1997; 014465.  
 PR 13-AUG-1996; US-023867.  
 PA (CHIR) CHIRON CORP.  
 PI Cohen F, Dubois-Stringfellow N, Dworki V, Innis MA,  
 PI Murphy JE, Tetsuo U, Zukermann R;  
 DR WPI: 98-139296/14.  
 PT Polycationic agents based on alpha-amino acids, able to complex  
 PT with nucleic acid - to facilitate its entry into cell, condense it  
 PT and protect it against serum degradation, particularly for use in  
 PT gene therapy  
 PS Disclosure: Page 77-80; 100pp; English.  
 CC This polynucleotide comprises the DNA sequence of vector plasmid  
 CC pCMVkmITR-EPI, which contains an Epstein-Barr virus (EBV) origin  
 CC of replication from plasmid pCEP4, a coding region for EBV nuclear  
 CC antigen A from pCEP4, a pair of inverted terminal repeats from  
 CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
 CC bovine growth hormone polyA sequence, and a kanamycin resistance  
 CC selectable marker. Polynucleotides encoding polypeptides, such as  
 CC erythropoietin or leptin, and ribozymes and antisense  
 CC polynucleotides can be inserted into the vector. The vector is  
 CC preferred for use in novel compositions and methods for improved  
 CC polynucleotide delivery into cells. In these methods, polycationic  
 CC agents are used to increase the frequency of uptake of a  
 CC nucleic acid (see also V21684-86) into a cell. The polycationic  
 CC agent can condense with the nucleic acid and inhibit serum and/or  
 CC nuclease degradation of the nucleic acid. The nucleic acid can be  
 CC a vector, may express a therapeutic protein or a vaccinating viral  
 CC or cancer antigen, or is itself therapeutic (antisense or  
 CC ribozyme). The methods and compositions can be used in the gene  
 CC therapy of many diseases.  
 SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T;

Query Match 2.6%; Score 57.2; DB 1; Length 9600;  
 Best Local Similarity 50.7%; Pred. No. 3.3e-05;  
 Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

OY 976 cggagactgttcatacagaccggagcgctgacagagcgcgacgctgagctgcag 1035  
 Db 713 CGGAGAGGCGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGAGGAGGCGG 772  
 OY 1036 cggcagagacttctcattcagaagcgctgagatgagtcacaaagatcccccagag 1095  
 Db 773 CAGGAGGCGGCGAGGAGGCGGCGAGGAGGAGGCGGCGAGGAGGAGGCGG 832  
 OY 1096 cagcagagagatggagcgcaagaagaaatggtccagaagcgacgagagaat 1155  
 Db 833 GGGCAGAGGCGGCGAGGCGGCGAGGAGGAGGCGGCGAGGAGGAGGCGG 892  
 OY 1156 gagagatacccggaagagatggaacagattgtgagagagagaagtttaagaagcaa 1215  
 Db 893 GAGGGGCGAGGAGGAGGAGGCGGCGAGGAGGCGGCGAGGAGGCGGAGGAGG 952  
 OY 1216 tgggaagaagactggggtcctcaagaagacag 1245  
 Db 953 CAGGAGCAGAGGAGGCGGCGAGGAGGCGGCGAG 982

RESULT 11

| ID | OS1731   |
|----|--|
| ID | Q51731 standard; DNA; 10396 BP.                                      |
| AC | Q51731;  |
| DT | 31-MAY-1994 (first entry)  |
| DE | Plasmid pCISBON for subcloning huHGF variants.                       |
| KW | Hepatocyte Growth Factor; HGF; variant; mutant; in vitro mutagenesis |
| KW | proteolysis resistant; liver; malignancy; CMV-driven;                |
| KW | Cyomegalovirus; episomal expression plasmid; ss.                     |
| OS | Synthetic.   |
| FH | Key  |
| FT | enhancer   |
| FT | Location/Qualifiers  |
| FT | 1..611   |
| FT | /tag= a  |
| FT | /note= "CMV enhancer/promoter"                                       |
| FT | 758..775   |
| FT | /tag= b  |
| FT | /label= SP6_promoter   |
| FT | 845..849   |
| FT | /tag= c  |
| FT | /note= "SP6 RNA start"   |
| FT | 902..966   |
| FT | /tag= d  |
| FT | /function= cloning_linker  |
| FT | 967..1107  |
| FT | /tag= e  |
| FT | /note= "SV40 poly A"   |
| FT | 1108..1531   |
| FT | /tag= f  |
| FT | /function= SV40_origIn   |
| FT | 1580..4189   |
| FT | /tag= g  |
| FT | /label= EBNA-1   |
| FT | 4190..6374   |
| FT | /tag= h  |
| FT | /function= oriP  |
| FT | 4295..4487   |
| FT | /tag= i  |
| FT | /note= "family of repeats"   |
| FT | 5866..5978   |
| FT | /tag= j  |
| FT | /note= "dyad region"   |
| FT | 6375..6457   |
| FT | /tag= k  |
| FT | /label= HSV-TK-terminator_3'-end                                     |
| FT | 6975..7975   |
| FT | /tag= l  |
| FT | /phenotype= neomycin resistance                                      |
| FT | /note= "h5 neomycin phosphotransferase gene"                         |
| FT | 7975..8112   |
| FT | /tag= m  |
| FT | /label= TK_Promoter  |
| FT | 8114..8594   |
| FT | /tag= n  |
| FT | /function= M13_Orl   |
| FT | 8595..10414  |
| FT | /tag= o  |
| FT | /label= delta_2a   |
| PN | WO9323541-A.   |
| PD | 25-NOV-1993.   |
| PF | 17-MAY-1993; U04648.   |
| PR | 18-MAY-1992; US-884811.  |
| PR | 18-MAY-1992; US-885971.  |
| PA | (GETH ) GENENTECH INC.   |
| PI | Godowski PJ, Lokker NA, Mark MR;                                     |
| PI | WPI: 93-386573/48.   |
| DR |  |
| PT | Hepatocyte growth factor variants - are resistant to proteolytic     |
| PT | cleavage into its two-chain form, used to treat malignancies         |
| PT | associated with HGF receptor   |
| PS | Example 1: FIG 6; 87pg: English.                                     |
| CC | Plasmid pCISBON (a pRK5 derivative) is an episomal CMV driven        |
| CC | expression plasmid. huHGF variants with enhanced receptor binding    |
| CC | activity were produced by site-directed mutagenesis. Stable          |
| CC | populations of preferred HGF variants were obtained by transfecting  |
| CC | human embryonic kidney 293 cells and then these were subcloned in    |

| CC | pc1SEBRON. See RS2940-R52949 for examples of pref. HGF variants.  |
|----|---|
| SQ | Sequence 10596 bp; 2625 A; 2571 C; 3024 G; 2376 T.  |
|    | Query Match 2.6%; Score 57.2; DB 1; Length 10596;<br>Best Local Similarity 50.7%; Pred. No. 3.5e-05;<br>Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps |
| QY | 976 cggagacgtcttcataagaaagaccggagacgagctctgcagaagacgagcagcgtctgagctcag 1035B  |
| Db | 2247 CGGGAGGGGGCAGGAGCAGGAGGGGGCAGAGACAGGAGGAGGGGCAGAGCAGGAGGAGGGG 2306B  |
| QY | 1036 cgcgcagagagcttctcatatgcagaagacgagctcgatgagatccacaagaatctccagag 1095B   |
| Db | 2307 CAGGAGGGGGCAGGAGGGGGCAGGAGGGGGCAGGACAGGAGGAGGGGCAGGACGACGAGCAG 2366B   |
| QY | 1096 cagcagagagatgtaggcgcgaagaagagagaagaattgcccagaagcgacagaggaatat 1155B  |
| Db | 2367 GGGCGAGGAGGGGCGAGGAGGGCGAGACAGGAGAGGGGCGAGAGCAGGAGGAGGCGCAG 2426B  |
| QY | 1156 gagaataccggaagagatctggaacagatctgtacgaagaggaagaatttaagaacaa 1215B   |
| Db | 2427 GAGGGCGCAGGAGCAGGAGGAGGAGGGGGCAGAGGGGCGAAGAGGGGCAGGAGGAGGAGG 2466B   |
| QY | 1216 tgggaagaagaacttggggtcctaagaagaacag 1245  |
| Db | 2487 CAGGAGCAGGAGGAGGGCGCAGGAGGGCGAG 2516   |

RESULT 12  
 T40348  
 AC T40348 standard; DNA; 10596 BP.  
 T40348;  
 DT 09-DEC-1996 (first entry)  
 DE Human; pcisEBON for expression of hepatocyte growth factor.  
 KW Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;  
 KW pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;  
 KW plasminogen; catalytic domain; serine protease; HGF variant;  
 KW HGF receptor; malignancy; chronic HGF receptor activation; ss.  
 OS Synthetic.  
 PN US5547856-A.  
 PD 20-AUG-1996.  
 PF 18-MAY-1992; 884811.  
 PR 18-MAY-1992; US-885971.  
 PR 18-MAY-1992; US-884811.  
 PR 13-JUL-1993; US-087783.  
 PA (GENTH ) GENENTECH INC.  
 PI Godowski PJ, Locker NA, Mark MR;  
 DR WPI; 96-392634/39.  
 PT New hepatocyte growth factor variants - are resistant to in vivo  
 PT proteolytic cleavage into a 2-chain form, useful as HGF antagonists  
 PS Example 1; Fig 6; 39pp; English.  
 CC This sequence represents the episomal CMV driven expression plasmid  
 CC pcisEBON which was used in the expression of variant human hepatocyte  
 CC growth factor (HGF). HGF is isolated from human serum and is a  
 CC disulphide linked heterodimer derived by proteolytic cleavage of the  
 CC pro-hormone between residues 494 and 495. This generates a molecule  
 CC composed of an alpha subunit of 440 amino acids (mol. wt. 69 kd) and  
 CC a beta subunit of 234 amino acids (mol. wt. 34 kd). The alpha and beta  
 CC subunits are encoded by a single open reading frame. The alpha and beta  
 CC contains four kringle domains based on their homology to kringle-like  
 CC domains in other proteins, e.g. prothrombin, plasminogen. The beta  
 CC subunit shows high homology to the catalytic domain of serine proteases.  
 CC However two of the three residues which form the catalytic triad of  
 CC serine proteases are not conserved in HGF. Therefore, the precise  
 CC function of the beta chain remains unknown. The invention includes HGF  
 CC variants which retain HGF receptor binding activity without having the  
 CC biological activity of wild-type HGF. They can be used for the treatment  
 CC of pathological conditions associated with the activation of a HGF  
 CC receptor such as malignancies associated with chronic HGF receptor  
 CC activation. The pcisEBON plasmid comprising the HGF coding sequence  
 CC may be used for manipulation of the HGF coding sequence and expression  
 CC of the variant HGF's of the invention.

|        |   |   |                    |           |               |         |
|--------|---|---|--------------------|-----------|---------------|---------|
| SQL    | Sequence  | 10596 BP;   | 2627 A;            | 2571 C;   | 3023 G;       | 2375 T; |
|        | Query Match   | 2.6%;   | Score 57.2;        | DB 1;     | Length 10596; |         |
|        | Best Local Similarity   | 50.7%;  | Pred. No. 3.5e-05; |           |               |         |
|        | Matches 137; Conservative   | 0;  | Mismatches 133;    | Indels 0; | Gaps 0;       |         |
| OY     | 976   | cggagacgtgcttatacacaaccgcggagcgctgtgcagagagcgcgacgctgtagctgcag  | 1035               |           |               |         |
| Db     | 2247  | CGGGGAGGGGCACAGACACAGAGGGGCGCAGAGCAAGAGAGAGGAGGCGCACAGAGCGAGG   | 2306               |           |               |         |
| OY     | 1036  | cggcaggagagcttcattatgcagaagcgcgctcgaggatgttgatgccaaagtctccaggag | 1095               |           |               |         |
| Db     | 2307  | CAGCAGGGGCCAAGAGAGGGCCAGAGGGGCGCAGAGCAAGAGAGGCGCACAGAGCGAGGAG   | 2366               |           |               |         |
| OY     | 1096  | cagcagagagatgagacgycgaagaagaaaagaatatgtcccagaagcgcaacagagaaat   | 1155               |           |               |         |
| Db     | 2367  | GGCGCAGAGAGGGGCAGAGAGGGGCGCAGAGCAAGAGAGGCGCACAGAGCGAGGCGCAG     | 2426               |           |               |         |
| OY     | 1156  | gagagatattccggaagagagatgtgaacagatgttagagaggaagagaagttaagaagcaa  | 1215               |           |               |         |
| Db     | 2427  | GAGGGGCGAGGACACAGAGAGGCGGCGCAGAGGGGCGCAGAGCGACGACGAGAGGCGG      | 2486               |           |               |         |
| OY     | 1216  | tgggaagaagactgggggtccaaggaacag                                  | 1245               |           |               |         |
| Db     | 2487  | CAGGAGCAAGAGAGGAGGGGCGACAGAGGGCGCAG                             | 2516               |           |               |         |
| RESULT | 13  |   |                    |           |               |         |
| ID     | X15650  |   |                    |           |               |         |
| AC     | X15650 standard; DNA; 10596 BP.   |   |                    |           |               |         |
| DC     | X15650:   |   |                    |           |               |         |
| DE     | 10-MAY-1999 (first entry)   |   |                    |           |               |         |
| KW     | Nucleotide sequence of plasmid PCIS.EBON.                                 |   |                    |           |               |         |
| RN     | Plasmid PCIS.EBON; hepatocyte growth factor; HGF; variant;                |   |                    |           |               |         |
| OS     | HGF receptor activation; ss.  |   |                    |           |               |         |
| PN     | Synthetic.  |   |                    |           |               |         |
| PP     | US5879910-A.  |   |                    |           |               |         |
| PD     | 09-MAR-1999.  |   |                    |           |               |         |
| PF     | 09-FEB-1994; 194087.  |   |                    |           |               |         |
| PR     | 09-FEB-1994; US-194087.   |   |                    |           |               |         |
| PA     | (GETH ) GENENTECH INC.  |   |                    |           |               |         |
| PI     | Godowski PJ, Lolkner NA, Mark MR;   |   |                    |           |               |         |
| DR     | WPI: 99-203949/17.  |   |                    |           |               |         |
| PT     | New hepatocyte growth factor variants - useful for studying               |   |                    |           |               |         |
| FT     | structure-function relationships in the wild type molecule and for        |   |                    |           |               |         |
| PT     | treating conditions associated with chronic hepatocyte growth factor      |   |                    |           |               |         |
| PS     | receptor activation   |   |                    |           |               |         |
| CS     | Example 1, Fig 6A-F; 40pp; English.                                       |   |                    |           |               |         |
| CC     | The present sequence represents the nucleotide sequence of PCIS.EBON,     |   |                    |           |               |         |
| CC     | which is used in the course of the invention. The specification           |   |                    |           |               |         |
| CC     | describes a hepatocyte growth factor (HGF) variant (HGfV) comprising an   |   |                    |           |               |         |
| CC     | amino acid (aa) alteration at or adjacent to position 692 of the          |   |                    |           |               |         |
| CC     | wild-type human HGF (huHGF) aa sequence. HGfV may be used in studies to   |   |                    |           |               |         |
| CC     | identify the structure-activity relationships of HGF in order to identify |   |                    |           |               |         |
| CC     | the functionally important domains in the aa sequence. It may also be     |   |                    |           |               |         |
| CC     | used to identify aa residues which are responsible for the interaction of |   |                    |           |               |         |
| CC     | HGF with its receptor, and those aa that are responsible for the          |   |                    |           |               |         |
| CC     | biological activity of it. Variants of HGfV which have enhanced receptor  |   |                    |           |               |         |
| CC     | binding affinity (compared to wild-type huHGF) and are more biologically  |   |                    |           |               |         |
| CC     | active than wild type huHGF, may be used as huHGF agonists. Conversely,   |   |                    |           |               |         |
| CC     | variants of HGfV which have enhanced receptor binding affinity (compared  |   |                    |           |               |         |
| CC     | to wild-type huHGF) but which are biologically inactive may be used as    |   |                    |           |               |         |
| CC     | huHGF antagonists, and may be used to block the binding of wild-type      |   |                    |           |               |         |
| CC     | huHGF to its receptor. This permits the treatment of pathogenic           |   |                    |           |               |         |
| CC     | conditions associated with the activation of an HGF receptor, such as     |   |                    |           |               |         |
| CC     | malignancies associated with chronic HGF receptor activation.             |   |                    |           |               |         |
| CC     | HGfV-immunoglobulin chimeras may be produced (by standard methods) and    |   |                    |           |               |         |
| CC     | used in protein A purification, immunohistochemistry, and                 |   |                    |           |               |         |
| CC     | immunoprecipitation techniques (in place of anti-HGF antibodies) or in    |   |                    |           |               |         |
| CC     | screening studies to identify inhibitors of HGF-HGF interactions.         |   |                    |           |               |         |
| SQ     | Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;                        |   |                    |           |               |         |

| Query Match | Similarity  | Score  | DB   | Length                           |
|-------------|---|--|------|----------------------------------|
| Best Local  | 50.7%   | Pred. No. 3, 5e-05   |      |                                  |
| Matches     | 137   | Conservative   | 0    | Mismatches 133; Indels 0; Gaps 0 |
| Qy          | 976   | cggagactcttcattgacagaccggagcggctgcgaagcgcgagcgttgactcag        | 1035 |                                  |
|             |   |  |      |                                  |
| Db          | 2247  | CGGGAGGGCGCAGAGACGACGAGGGGGCAGAGACAGAGAGGGGCGAGACGAGAGAGGGC    | 2306 |                                  |
| Qy          | 1036  | cggcaggaagctcttcattgacagacggctgcgactgagatcccaagatctccagag      | 1095 |                                  |
|             |   |  |      |                                  |
| Db          | 2307  | CAGGAGGGCGCAGAGAGGGGCGCAGAGGGGCGCAGAGAGAGAGAGGGGCGCAGAGAGGAG   | 2366 |                                  |
| Qy          | 1096  | cagcaggaagatgtgagcggcgaagagaaagaattgtcccaagagcagcagagaaat      | 1155 |                                  |
|             |   |  |      |                                  |
| Db          | 2367  | GGCGCAGGAGGGCGCAGAGGGGCGCAGAGAGAGGGGCGCAGAGAGAGGGGCGCAG        | 2426 |                                  |
| Qy          | 1156  | gagagataccgaggaagatgtgacacgattgtgagagaggaagaaagttaagaacaa      | 1215 |                                  |
|             |   |  |      |                                  |
| Db          | 2427  | GAGGGCGCAGGAGCAGGAGAGAGGGGCGCAGAGGGGCGCAGAGAGGGGCGCAGAGAGAGGGG | 2486 |                                  |
| Qy          | 1216  | tgggaagaagacttggggtccaagaagacag                                | 1245 |                                  |
|             |   |  |      |                                  |
| Db          | 2487  | CAGGACGACGAGGAGGGGCGCAGAGGGGCGCAG                              | 2516 |                                  |
| RESULT      | 14  |  |      |                                  |
| V55830      |   |  |      |                                  |
| ID          | V55830 standard; DNA; 795 BP.   |  |      |                                  |
| AC          | V55830;   |  |      |                                  |
| DT          | 18-NOV-1998 (first entry)   |  |      |                                  |
| DE          | FLGA insert stabilising polypeptide encoding DNA.                         |  |      |                                  |
| KM          | Fusion protein; stabilising polypeptide; proteolytic degradation;         |  |      |                                  |
| KW          | Resistance; half-life; autoimmune disease; inflammation; nitro drug;      |  |      |                                  |
| KM          | Ikappab regulator protein; inflammatory bowel disease; in vivo imaging;   |  |      |                                  |
| KW          | nitroreductase protein; enzyme therapy; prodrg therapy; protease;         |  |      |                                  |
| KM          | cancer; pathological condition; ss.                                       |  |      |                                  |
| OS          | Epstein-barr virus.   |  |      |                                  |
| FT          | Key   |  |      |                                  |
| FT          | Location/Qualifiers   |  |      |                                  |
| FT          | 1..788  |  |      |                                  |
| FT          | CDS   |  |      |                                  |
| FT          | /*tag= a  |  |      |                                  |
| FT          | /product= "stabilising polypeptide"                                       |  |      |                                  |
| PN          | WO9822577-A1.   |  |      |                                  |
| PD          | 28-MAY-1998.  |  |      |                                  |
| PF          | 17-NOV-1997; IB1508.  |  |      |                                  |
| PR          | 25-JUN-1997; US-048945.   |  |      |                                  |
| PR          | 15-NOV-1996; US-030986.   |  |      |                                  |
| PA          | (MASU/) MASUCCI M. G.   |  |      |                                  |
| PI          | Masucci MG;   |  |      |                                  |
| DR          | WPI: 98-312463/27.  |  |      |                                  |
| DR          | P-PSDB: W79128.   |  |      |                                  |
| PT          | New fusion proteins resistant to proteolytic degradation -                |  |      |                                  |
| PT          | comprising a core protein with a stabilising polypeptide comprising       |  |      |                                  |
| PS          | a peptide sequence containing glycine repeats                             |  |      |                                  |
| PS          | Disclosure; Fig 3; 120pp; English.  |  |      |                                  |
| CC          | This DNA encodes a stabilising polypeptide and is the FLGA insert of the  |  |      |                                  |
| CC          | invention. The invention provides a method for increasing the resistance  |  |      |                                  |
| CC          | of a core protein to proteolytic degradation that comprises linking or    |  |      |                                  |
| CC          | inserting onto or into the core protein a stabilising polypeptide of      |  |      |                                  |
| CC          | formula [(Glya)X(Glyb)Y(Glyc)Z]n where Glya, Glyb, Glyc are 1-6           |  |      |                                  |
| CC          | sequential Gln residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,     |  |      |                                  |
| CC          | Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not   |  |      |                                  |
| CC          | be identical from n repeat to n repeat. Alternatively a nucleic acid      |  |      |                                  |
| CC          | encoding the stabilising polypeptide can be linked onto or inserted into  |  |      |                                  |
| CC          | a nucleic acid encoding a core protein. The fusion proteins of the        |  |      |                                  |
| CC          | invention are more resistant to degradation by proteases and, thus, have  |  |      |                                  |
| CC          | a longer half-life than the unfused core protein. The products can be     |  |      |                                  |
| CC          | used for treating autoimmune diseases, cancer and inflammation. In        |  |      |                                  |
| CC          | particular, the core protein may be an Ikappab regulator protein for the  |  |      |                                  |
| CC          | treatment of inflammatory bowel disease, or a nitroreductase protein      |  |      |                                  |
| CC          | which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  |  |      |                                  |
| CC          | or other pathological conditions. The fusion proteins can also be used in |  |      |                                  |
| CC          | diagnostic methods such as in vivo imaging.                               |  |      |                                  |



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 22:33:48 ; Search time 3256.04 Seconds  
(without alignments)  
2691.327 Million cell updates/sec

Title: US-09-502-945-5  
Perfect score: 2162  
Sequence: 1 ccctggcccgctgcgctgcg.....tatttcagcttaaaaaa 2162

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202661650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : EST:\*

|     |            |
|-----|------------|
| 1:  | em_est1:*  |
| 2:  | em_est2:*  |
| 3:  | em_est3:*  |
| 4:  | em_est4:*  |
| 5:  | em_est5:*  |
| 6:  | em_est6:*  |
| 7:  | em_est7:*  |
| 8:  | em_est8:*  |
| 9:  | em_est9:*  |
| 10: | em_est10:* |
| 11: | em_est11:* |
| 12: | em_est12:* |
| 13: | em_est13:* |
| 14: | em_est14:* |
| 15: | em_est15:* |
| 16: | em_est16:* |
| 17: | em_est17:* |
| 18: | em_est18:* |
| 19: | em_est19:* |
| 20: | gb_est1:*  |
| 21: | gb_est2:*  |
| 22: | gb_est3:*  |
| 23: | gb_est4:*  |
| 24: | gb_est5:*  |
| 25: | gb_est6:*  |
| 26: | gb_est7:*  |
| 27: | gb_est8:*  |
| 28: | gb_est9:*  |
| 29: | gb_est10:* |
| 30: | gb_est11:* |
| 31: | gb_est12:* |
| 32: | gb_est13:* |
| 33: | gb_est14:* |
| 34: | gb_est15:* |
| 35: | gb_est16:* |
| 36: | gb_est17:* |
| 37: | gb_est18:* |
| 38: | gb_est19:* |
| 39: | gb_est20:* |
| 40: | gb_est21:* |
| 41: | gb_est22:* |
| 42: | gb_est23:* |
| 43: | gb_est24:* |
| 44: | gb_est25:* |

|      |            |
|------|------------|
| 45:  | gb_est26:* |
| 46:  | gb_est27:* |
| 47:  | gb_est28:* |
| 48:  | gb_est29:* |
| 49:  | gb_est30:* |
| 50:  | gb_est31:* |
| 51:  | gb_est32:* |
| 52:  | em_est20:* |
| 53:  | em_est21:* |
| 54:  | em_est22:* |
| 55:  | em_est23:* |
| 56:  | em_est24:* |
| 57:  | em_est25:* |
| 58:  | em_est26:* |
| 59:  | gb_est33:* |
| 60:  | gb_est34:* |
| 61:  | gb_est35:* |
| 62:  | gb_est36:* |
| 63:  | gb_est37:* |
| 64:  | gb_est38:* |
| 65:  | em_est27:* |
| 66:  | em_est28:* |
| 67:  | em_est29:* |
| 68:  | em_est30:* |
| 69:  | gb_est39:* |
| 70:  | gb_est40:* |
| 71:  | gb_est41:* |
| 72:  | gb_est42:* |
| 73:  | gb_est43:* |
| 74:  | gb_est44:* |
| 75:  | em_est31:* |
| 76:  | em_est32:* |
| 77:  | em_est33:* |
| 78:  | em_est34:* |
| 79:  | gb_est45:* |
| 80:  | gb_est46:* |
| 81:  | gb_est47:* |
| 82:  | gb_gss1:*  |
| 83:  | gb_gss2:*  |
| 84:  | gb_gss3:*  |
| 85:  | gb_gss4:*  |
| 86:  | em_gss1:*  |
| 87:  | em_gss2:*  |
| 88:  | em_gss3:*  |
| 89:  | em_gss4:*  |
| 90:  | gb_gss5:*  |
| 91:  | gb_gss6:*  |
| 92:  | gb_gss7:*  |
| 93:  | gb_gss8:*  |
| 94:  | gb_gss9:*  |
| 95:  | em_gss5:*  |
| 96:  | em_gss6:*  |
| 97:  | em_gss7:*  |
| 98:  | em_gss8:*  |
| 99:  | em_gss9:*  |
| 100: | em_gss10:* |
| 101: | em_gss11:* |
| 102: | gb_gss10:* |
| 103: | gb_gss11:* |
| 104: | em_gss12:* |
| 105: | gb_gss12:* |
| 106: | gb_gss13:* |
| 107: | gb_gss14:* |
| 108: | gb_gss15:* |
| 109: | gb_gss16:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

|        |   |
|--------|---|
| Result | * |
| Query  |   |

| No.  | Score | Match | length | DB  | ID       | Description |
|------|-------|-------|--------|-----|----------|-------------|
| C 1  | 578.2 | 26.7  | 516    | 61  | AL045824 | DKFZP434H   |
| C 2  | 477.2 | 22.1  | 512    | 79  | AM300030 | AM300030    |
| C 3  | 457.8 | 21.2  | 473    | 44  | AI126960 | qk18d11.x   |
| C 4  | 440.6 | 20.4  | 485    | 39  | AA680926 | ak22d06.x   |
| C 5  | 437.6 | 20.2  | 467    | 51  | AI761354 | w160e08.x   |
| C 6  | 435.4 | 20.1  | 464    | 51  | AI739238 | w114d07.x   |
| C 7  | 423   | 19.6  | 473    | 63  | AI984995 | wr89d06.x   |
| C 8  | 405.6 | 18.8  | 636    | 26  | w221d60  | 63d6 Human  |
| C 9  | 403   | 18.6  | 411    | 49  | AI628885 | ly79c02.x   |
| C 10 | 403   | 18.6  | 411    | 49  | AI629007 | ly79e02.x   |
| C 11 | 394.2 | 18.2  | 407    | 63  | AI989344 | w227e03.x   |
| C 12 | 372.8 | 17.2  | 384    | 60  | AI796741 | wai4d02.x   |
| C 13 | 371.2 | 17.2  | 385    | 44  | AI272081 | qj78f10.x   |
| C 14 | 371.2 | 17.2  | 465    | 80  | AM390711 | RC1-ST019   |
| C 15 | 369.2 | 17.1  | 386    | 46  | AI453367 | tj37c11.x   |
| C 16 | 355.2 | 16.4  | 360    | 63  | AI199149 | ws10b04.x   |
| C 17 | 351.4 | 15.8  | 367    | 79  | AM304132 | xs13f01.x   |
| C 18 | 341.4 | 15.5  | 381    | 80  | AM351654 | QV2-CT014   |
| C 19 | 335.6 | 15.5  | 342    | 51  | AI738967 | w112f02.x   |
| C 20 | 334   | 15.4  | 342    | 51  | AI738958 | w112e03.x   |
| C 21 | 304.6 | 14.1  | 311    | 35  | AA551799 | nk04d11.s   |
| C 22 | 292.2 | 13.5  | 317    | 32  | AA344913 | EST50856    |
| C 23 | 290.4 | 13.4  | 335    | 78  | AA121174 | z188b08.s   |
| C 24 | 287.8 | 13.3  | 316    | 74  | AA192301 | x184d12.x   |
| C 25 | 270.4 | 12.5  | 272    | 44  | AI270726 | qf93d07.x   |
| C 26 | 267.4 | 12.4  | 289    | 32  | AA348243 | EST54707    |
| C 27 | 266.4 | 12.3  | 278    | 63  | AM013904 | UI-H-B10-   |
| C 28 | 248.8 | 11.5  | 279    | 61  | AI869919 | we75h03.x   |
| C 29 | 240.8 | 11.1  | 244    | 61  | AI783811 | tu79c11.x   |
| C 30 | 197.4 | 9.1   | 216    | 61  | AI869907 | we75b03.x   |
| C 31 | 180.8 | 8.4   | 206    | 31  | AA327309 | EST30621    |
| C 32 | 160   | 7.4   | 204    | 28  | AA121198 | z188b08.r   |
| C 33 | 143   | 6.6   | 353    | 49  | AI642234 | v68c05.y    |
| C 34 | 131.2 | 6.1   | 221    | 64  | AM062449 | MRO-CT006   |
| C 35 | 120.6 | 5.6   | 473    | 36  | AA616807 | v68c05.r    |
| C 36 | 105   | 4.9   | 463    | 105 | AQ030634 | RPCT11-9-   |
| C 37 | 68.2  | 3.2   | 823    | 84  | B03396   | B03396      |
| C 38 | 67.2  | 3.1   | 443    | 102 | AQ100902 | HS_-3060-   |
| C 39 | 58.2  | 2.7   | 997    | 82  | CNS0507E | AL060767    |
| C 40 | 57.2  | 2.6   | 587    | 86  | C79507   | drosophil   |
| C 41 | 56.6  | 2.6   | 571    | 92  | AO911910 | C79507      |
| C 42 | 55.8  | 2.6   | 476    | 36  | C79500   | MOUS        |
| C 43 | 55.6  | 2.6   | 445    | 34  | AA64787  | x83g10.r    |
| C 44 | 55.2  | 2.6   | 498    | 50  | AI676456 | ecmeST021   |
| C 45 | 54.6  | 2.5   | 657    | 80  | AM369742 | IL0-BT015   |

## ALIGNMENTS

| Description          |
|----------------------|
| AL065824 DKFZ434H3   |
| AM300030 X55501.x    |
| AI269560 qx18d11.x   |
| AA6806296 skk2206.o  |
| AI761354 w160e08.x   |
| AI792328 w114d07.x   |
| AI984995 w8r9e06.x   |
| W22160_6336 Human    |
| AI628895 LY79e02.x   |
| AI629907 LY79e02.x   |
| AI989344 w212e03.x   |
| AI976741 wa14d02.x   |
| AI272081 q7f8t10.x   |
| AM300711 RC1-EST019  |
| AI453367 t337c11.x   |
| AI919142 xs10b04.x   |
| AM304132 xs13f01.x   |
| AM331654 OV2-EST0144 |
| AI738967 w12f202.x   |
| AI738958 w12e03.s    |
| AA551799 nk04d11.s   |
| AA344913 EST50856    |
| AA121174 z189b08.x   |
| AA192301 x184d12.x   |
| AI210726 q7g3c07.x   |
| AA348243 EST54707    |
| AM013904 U1-H-B10-1  |
| AI669919 w675b03.x   |
| AI783811 U79e11.x    |
| AI669307 w675b03.x   |
| AA327309 EST30621    |
| AA121158 z189b08.x   |
| AI642234 y198e05.y   |
| AA662449 MR-CT0060   |
| AA616807 w68e05.o    |
| AO320634 RPT111-9    |
| B03396 GSTR-178D10   |
| AO100902 HS_-3060    |
| AI606767 drosophili  |
| CT9507 C79507 Mous   |
| AG5019110 RPCI_-23-2 |
| C79500 C79500 Mous   |
| AA64787 zx83g10.r    |
| AI676456 eBT015      |
| AM369742 IL0-T015    |

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Euthelia; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 616)  
AUTHORS Koehrer, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S  
TITLE EST (Koehrer, et al.)  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced gi:3137188.

Am Klopferspitz 18a D-82152 Martinsried, Germany  
This is the 3' sequence of the clone insert

```

FEATURES
source
    Clone from S. Wiemann, Moleculare Genome Analysis, German Cancer
    Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
    sequenced by BKFZ within the cDNA sequencing consortium of the
    German Genome Project.
    r1 sequence also available.
    This clone is available at the RZPD in Berlin.
    Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14055
    Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
    Location/Qualifiers
        1..616
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKFZP434H236"
            /clone_1lb="434 (synonym: htes3)"
            /tissue_type="testis"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
134 a 142 c 105 g 155 t

```

|                       |              |                    |              |            |
|-----------------------|--------------|--------------------|--------------|------------|
| Query Match%          | 26.7%        | Score 578.2        | DB 61        | Length 616 |
| Best Local Similarity | 98.4%        | Pred. No. 2,1e-136 |              |            |
| Matches 605           | Conservative | 0                  | Mismatches 8 | Indels 2   |
|                       |              |                    |              | Gaps 2     |

|            |      |   |      |
|------------|------|---|------|
| QY         | 1529 | atgcaactcaagcgcaagatgtgtagaagacttaacccctgctggaagctgaagctgcgctg    | 1588 |
| Db         | 613  | ATGCGAATTCAAAGCGCAAGATTGTGACAGACTACACCCCTGGCTGAGGCTGACCGTCCCTG    | 554  |
| QY         | 1589 | cagaagccttgaatcaaggcgggagacttgatcgaccttgtgtgtgcgtctgtcccca        | 1648 |
| Db         | 553  | CAGAAGCCCTGGAAATCAGGCGGGGAGCTGGATGACCTTGTGTGTCCTGCGCCGCCCA        | 494  |
| QY         | 1649 | aagaagatgatacatgagctgcgagactctctgtgcgtgaatgccaaaagggaaaccaaatctac | 1708 |
| Db         | 493  | AAGAGATGATGACGATGAGCTGACCTTCTTCTTCTGAAGTCCAAAAGGGGAAACCAATTATAC   | 435  |
| QY         | 1709 | gcgttaagaaacagtgagctgcgggcccaactcgtgaacacaaagccttgcagacactt       | 1768 |
| Db         | 434  | -CGTTAGGAAGAACAGTGAGCTCGGGCCCAACCTGTCGAAACAAAGCCTGCGATCAGCCTT     | 376  |
| QY         | 1769 | gaaggaagcacaatgacacacacacaagatgcatcctcttgagactgaatctatcacacagg    | 1828 |
| Db         | 375  | GAGAGAGGCGACACTACACACACAGATGGCATCTTGGAGCCTGAATCTATCACCCAGG        | 316  |
| QY         | 1829 | aatctcaactcctcttgcggcccgaaacccagggccaatagaagaaacagctgcggccaactt   | 1888 |
| Db         | 315  | AATCTCAAACTCCTTTGGCCCTGAAACCAAGGGCGAATATAGGAACAGGCTCGGGCCACTCT    | 256  |
| QY         | 1889 | tttgaagggccaatgttggaggaagaaaggagacgccagcgttttgggaaagaatctcaagat   | 1948 |
| Db         | 255  | TCTGAAAGGCCAACGTGGAGGAAAGGAGAGAGCAGCGCAATTTGGAGGAAGATCTCAAGAT     | 196  |
| QY         | 1949 | ccgaagcctatctcctctctctgcgcgaagtgaatttgtctctccagcttgcggggagc       | 2008 |
| Db         | 195  | CCGAGCTCTATTCCTTTCTCTCTGCGGCCAGTGAATTTGGTCTCTGCCAGCTGTGGGGAC      | 136  |
| QY         | 2009 | tccctcttgaaccctaataagaagcccaacttgaagtcctctctctctccalcctctcctt     | 2068 |
| Db         | 135  | TCTCTTCCTTAAACCTTAATTAAGACCCCACTGGAGAGTCTCTCTCTCCATCTCCCTACCT     | 76   |
| QY         | 2069 | gcacctgcgctctaattgctgcgcgaagtttgaactccaacacttactctgaagctaatat     | 2128 |
| Db         | 75   | GCCCTCTGCTCTAATTGCTGCCAGGATGTGACTCCAAACCTTACTCTGAGCTCATTTAAT      | 16   |
| QY         | 2129 | aaataaacagattt 2143   |      |
| Db         | 15   | AAATAAACAAGATTT 1   |      |
| RESULT     | 2    |   |      |
| FW300030/c |      |   |      |



LOCUS AM300030 512 bp mRNA EST 18-JAN-2000  
DEFINITION xs55h01.x1 NCI\_CGAP\_Kid11 Homo sapiens cDNA clone IMAGE:2773585 3',  
mRNA sequence.  
ACCESSION AM300030  
VERSION AM300030.1 GI:6709707  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 512)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3137941.  
CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbrr/image/image.html](http://www.bio.llnl.gov/dbrr/image/image.html)

FEATURES  
source  
Seq primer: -40UP from gibco  
High quality sequence stop: 462.  
Location/Qualifiers  
1..512  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2773585"  
/clone\_1lb="NCI\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI\_CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502835). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 117 a 105 c 158 g 132 t  
ORIGIN

Query Match 22.1%; Score 477.2; DB 79; Length 512;  
Best Local Similarity 98.1%; Pred. No. 8.9e-111;  
Matches 504; Conservative 0; Mismatches 8; Indels 2; Gaps 2;

QY 1638 tcggcccccgaagagatgacatgagctgacgtcttctgctgaagtccaaaaggga 1697  
|||||  
DB 512 TCTGCCCCCAAGAGATGAGCAGATGAGCTGACTTCTT-CTGAAGTCCAAAAGGGAA 454  
QY 1698 accaaattcagcgtttagaagaacagtgagctcggcccccacgtcgtgaacaaagcctc 1757  
|||||  
DB 453 ACCAAATTCAC -CGTTAGGAAACAGTGAAGTCCGGCCGCCACCTCGTGAACACAAAGCCTC 395  
QY 1758 ggaacagccttgagagagcacaacacacacagatggcatcctctggacctgaatc 1817  
|||||  
DB 394 GGAATCAGCTTGAAGAGCGCACACACACACAGATGACATCTTGAGACCTGAATC 335  
QY 1818 taccaccagggaatccaaactcccttgcccttgaaacaggccagatagaagaagct 1877  
|||||  
DB 334 TATCACCACGAGAACTCAAACTCCCTTGGCCCTGAACCGGGCCAGATTAAGAACAGCT 275  
QY 1878 cgggcaacttttgaaggccaatgtgagaaaggagagcagccgcttggagaag 1937  
|||||

DB 274 CGGGCCACTCTTCTGAAGGCCAACGTGAGAAAGGAGAGCAGCCATTGGAGAG 215  
LOCUS  
QY 1938 attcaagatccagactcctcattccttctctgcccagtgatattgctctccag 1997  
|||||  
DB 214 ATCTCAAGGATCCAGACTCTCATCTCTTCTGCGCCAGTAATTTGGCTCTCCAG 155  
QY 1998 cttggggagactccttctctgaacccttaataagaccacatggagctctctccat 2057  
|||||  
DB 154 CTCTGGGGAGACTCTTCTTGAACCTTAATAAACCCTGAGTCTCTCTTCAT 95  
QY 2058 ccctctctctgcccctctgctcctaattgctgcagagattgctacccaacttactctg 2117  
|||||  
DB 94 CCTCTCTCTGCGCCCTGCTGCTTAATTCGCGCAGAGTTTCACCTCAACCTTACTCG 35  
QY 2118 agctcattataaataacagatttcca 2151  
|||||  
DB 34 AGCTCATTAATAAATAAACGATTAATTTCCA 1

RESULT 3  
AI269960/c 473 bp mRNA EST 28-JAN-1999  
LOCUS  
DEFINITION qk1bd11.x1 NCI\_CGAP\_Kid3 Homo sapiens cDNA clone IMAGE:1869333 3',  
similar to TR:Q14160 Q14160 KIAA0147 PROTEIN ;, mRNA sequence.  
ACCESSION AI269960  
VERSION AI269960.1 GI:3889127  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 473)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbrr/image/image.html](http://www.bio.llnl.gov/dbrr/image/image.html)

Insert length: 520 Std Error: 0.00  
Seq primer: -40UP from gibco  
High quality sequence stop: 451.  
Location/Qualifiers  
1..473  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1869333"  
/clone\_1lb="NCI\_CGAP\_Kid3"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer,  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not  
I and Eco RI sites of the modified pT73 vector. mRNA  
source: 2 pooled kidneys. Library went through one round  
of normalization. Library constructed by Bento Soares and  
M. Fatima Bonaldo."

BASE COUNT 92 a 159 c 118 g 104 t  
ORIGIN

Query Match 21.2%; Score 457.8; DB 44; Length 473;  
Best Local Similarity 99.6%; Pred. No. 7.3e-106;

Matches 459; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy 654 tcaatttctgtcgaatcctggggcgctgagagcagcctggctccctggaaatcggga 713
    |||
Db 461 TCAGTTGTGTGGAAATCGGGGGCGTGGAGCAGCAGCCTGGCTCCCGGAAATCGGGA 402
Qy 714 aaacaagagaagaagcttctatcagccttgtaagctcccaagagccttgctcagcat 773
    |||
Db 401 AAACAAGAGAAAGAGCTTCATCAGCCTGTAGGCTCCGAGGCTTGCTGGTCAGCAT 342
Qy 774 ttccagggccccatccagaagcctggcatcttcatcagcatgtgaaacctgctccct 833
    |||
Db 341 TTCCAGCGGGCCCATCCAGAAAGCTGGCATCTTTATCACCCATGTGAACCTGGCTCCCT 282
Qy 834 gtctcgtagtggtgattggagataggaggaaccagattgtcgaatcgaatgagctgact 893
    |||
Db 281 GTCTCTGAGGTGGATTTGGAGATAGGGGACCAATTTGTCGAATGATAGCTGCACTT 222
Qy 894 ctctaactggaatcaaaagagcgttaaatgtctgaaataaagcgcagcctgacct 953
    |||
Db 221 CTCTAACCTGATCACAAGAGAGCTGTAAATGTGCTGAAGATAGCCGAGCTGACCAT 162
Qy 954 ctccatttgatcgtcagctgctggcgagctgtctcatgacagacgggagcgctggcaga 1013
    |||
Db 161 CTCCATTTGATGCTGACGTGGCGGGAGCTGTTCATGACAGACCGGAGCGCTGGCAGA 102
Qy 1014 ggcggcgagcgtgagctgcagcagcgagagcttctatcagaaagcgctggcgatga 1073
    |||
Db 101 GCGCGGGCAGCTGAGCTGACGCGGAGAGAGCTTCTCATGCAAGAGCGGCTGGCAATGA 42
Qy 1074 gtccaacaagatcctccagagcagcagagatgagcggc 1114
    |||
Db 41 GTCCAACAAGATCTCCAGAGACAGCAGAGATGAGCGGC 1
RESULT 4
AA860926/c 485 bp mRNA EST 04-JAN-1999
DEFINITION ak22d06.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1406699
ACCESSION AA860926
VERSION AA860926.1 GI:2953066
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 485)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On Jan 19, 1998 this sequence version replaced gi:2287585.
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMIL at:
www-bio.llnl.gov/bdrp/image/image.html
FEATURES
source 1..485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1406699"
/location="Soares_testis_NHT"
```

```
/sex="male"
/lab_host="DH10B"
/notice="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTTCACATCTGAAGTGGAGCGGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization to Cots, and was
constructed by Bento Soares and M. Fatima Bonaldo."
```

```
Query Match 20.4%; Score 440.4; DB 39; Length 485;
Best Local Similarity 98.1%; Pred. No. 2e-101;
Matches 477; Conservative 0; Mismatches 6; Indels 3; Gaps 3;
Qy 1666 gctgaccttctgtcgtgaagtcctcaaaagggaacaaattcagcgcttagaagaacagcga 1725
    |||
Db 485 GCTGACCTTCTT-CTGAAGTCCAAAAGGGGAACCAAAATTCAC-CGTTAGGAAACAGTGA 428
Qy 1726 gtccggccccacctcgtggaacaaagcctcggaaccaagccttgagagagcgacatgac 1785
    |||
Db 427 GCTCGGGCCCCACCTCGTGAACACAAAGCCTTGAGACAGCTTGAGAGAGCGCACACTAC 368
Qy 1786 acacacagatgacatccttggagctgaatctatcaccaggaatctcaactccctt 1845
    |||
Db 367 ACACACAGAGTGGATCTCTGGAGCTGAATCTACCCAGAGATCTCAAACTCCCTTT 308
Qy 1846 ggccttgaaacagagcagataagaaagcctcgggccactttttggaagcgaatggg 1905
    |||
Db 307 AG-CCTGAACCAAGGCGCCAGATPAGGAACAGCTGGGGCCACTCTTGTGAAGGCCAATGTGG 249
Qy 1906 aggaagaggaagcagcagccgtttggagaagatctcaagatccagactcactcctt 1965
    |||
Db 248 AGGAAGAGGAGACGACGACCGCTTTGGAGAGATCTCAAGATCTCAGACTCATTTCTT 189
Qy 1966 tcccttgcccagtgaaattgctctccacagcttggggagctccttcccttgaacctta 2025
    |||
Db 188 TCCTCTGGCCAGTGAATTTGGTCTCTCCAGCTCTGGGGGAGACCTCTTGAAACCTTA 129
Qy 2026 ataagaacccactgagctctctctccatccctcctcctcctcctcctcctcaattg 2085
    |||
Db 128 ATAAAGACCCCACTGAGCTCTCTCCATCCCTCTCTGCTCTGCTCTTAATTG 69
Qy 2086 ctggcagagctgacatccaaaccttactctgagctcatataaataaacaagattat 2145
    |||
Db 68 CTGGCAGATTTCTCCTCAAACTTACTCTGAGCTCAATTAATAAATAACAGATTAT 9
Qy 2146 ttcca 2151
    |||
Db 8 TTTCGA 3
RESULT 5
A1761354/c 467 bp mRNA EST 24-JUN-1999
LOCUS A1761354
DEFINITION w160e08.x1 NCI_CGAP_Col16 Homo sapiens cDNA clone IMAGE:2394662 3',
mRNA sequence.
ACCESSION A1761354
VERSION A1761354.1 GI:5177021
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 467)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
```

JOURNAL  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3138298.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

FEATURES  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 467

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 108 a 96 c 143 g 120 t  
ORIGIN

Query Match 20.2%; Score 437.6; DB 51; Length 467;  
Best Local Similarity 98.7%; Pred. No. 9, 9e-101;  
Matches 462; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

1691 aggggaaccaatttcacgcttaggaagaagtcgcgcgcacccctcgtgaacaca 1750  
|||||  
467 AGGGGAACCAATTTCAC -CGTTAGCAAAACAGTACGCTCCGCCCACTCTGAACACA 409  
1751 aagcctcgagccagccttggagagagccacatgacacacacagatgcatccttggagc 1810  
|||||  
408 AAGCCTCGAGCCAGCCTTGAGAGGCCACACACACAGATGCATCTGCGGAC 349  
1811 ctgaatctatcacccaggaaatcacaatcccttggccttgaaccagggccagataag 1870  
|||||  
348 CTGAATCTATCACCCAGGAAATCTCAACTCTTGGCCCTGAACCGGGCCAGATTAAG 289  
1871 aacagctcgagccact-ttttgaaggccaattggagagaaaggagccagccgtt 1929  
|||||  
288 AACAGCTGGGCACTTTTGAAGCCAAATGGAGAAAGGAGACCAACCCGCTT 229  
1930 gggagaagatctcaagatccagactctccttctccttggccaggaatttggc 1989  
|||||  
228 GGGAGAAGATCTCAAGATCCAGACTCTCATTCCTTCCCTGGCCAGTAATTTGTC 169  
1990 tctccagcttggggagactccttcttgaaccctaagaaccacatgagctctc 2049  
|||||  
168 TCTCCAGCTCTGGGGACCTCTTCCCTTGAACCTAATAAGACCCCACTGAGCTCTCT 109  
2050 ctctcactcctctcctcctcctctgcttaatttggcagagatgttaaccacaac 2109  
|||||  
108 CTCTCACTCCCTCTCTCTGCTGCTCTAATGCTGCAAGATGTCTCACTCAAC 49  
2110 ttactctgagctcataataaataaacaagattatttccagcttaa 2157  
48 TTACTGTGAGCTCATTAATAAATAAACAGATTATTTTCCAGCTAAA 1

RESULT 6  
LOCUS A1739238/c  
DEFINITION w14g07.x1 NCI-CGAP\_Col6 Homo sapiens CDNA clone IMAGE:2390268 3',  
mRNA sequence.  
ACCESSION A1739238  
VERSION A1739238.1 GI:5101219  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 464)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jun 5, 1998 this sequence version replaced gi:3189064.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,  
M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 582 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 421.

FEATURES  
source

1. 464  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1ib="NCI-CGAP\_Col6"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pT73D-Pac (pharmacia) with a  
modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Col6 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1057416-1061255, and 1144584-1145351).  
Subtraction by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 104 a 90 c 140 g 130 t  
ORIGIN

Query Match 20.1%; Score 435.4; DB 51; Length 464;  
Best Local Similarity 98.7%; Pred. No. 3, 6e-100;  
Matches 439; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

1718 aacatgagagcttcggcccccctcgtgaacacaaagctcgagccagccttggagagc 1777  
|||||  
464 AACAGTAGGCTCCGGCCCACTCGTAACAAAGCCTCGACCAAGCTTGAAGAGGC 405  
1778 cacatgacacacacacagatgcatccttggagactgatatcaccagaatctcaaa 1837  
|||||  
404 CACACTACACACACACATATGATCTTGGAGCTGATATATCACCCAGAAATCTCAA 345  
1838 ctcccttggccctgaacacagggccagataaagaacagctcgggccaatttgaagc 1897  
|||||  
344 CTCCTTTGGCCCTGAACCGAGGCGCAGATTAAGGAACAGCTCGGGCCACTCTTGAAGGC 285  
1898 caatgaggaggaaggagagcagccgcttggggaagatctcaagatcagagctc 1957  
|||||

[illegible]

|            |   |
|------------|---|
| RESULT     | 7   |
| A1984995/c |   |
| LOCUS      | A1984995      473 bp    mRNA                  EST         31-AUG-1999 |
| DEFINITION | wf99g66.x1 NCI_CGAP_Kid1 Homo sapiens cDNA clone IMAGE:294906 3' ,    |
| ACCESSION  | A1984995<br>A1984995    mRNA sequence.                                |
| VERSION    | A1984995.1 GI:5812272   |
| KEYWORDS   | EST.  |
| SOURCE     | human.  |
| ORGANISM   | Homo sapiens  |

| REFERENCE          | AUTHORS                    | TITLE   |
|--------------------|----------------------------|---|
| 1 (pages 1 to 473) | NCI-CCAP                   | <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . |
|                    | National Cancer Institute, | Cancer Genome Anatomy Project (CGAP),   |

JOURNAL COMMENT  
Unpublished (1997)  
On May 18, 1998 this sequence version replaced gi:3138171.  
Contact: Robert Strausberg, Ph.D.

Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILM at:  
[www-bio.1ni.gov/dbfp/image/image.html](http://www-bio.1ni.gov/dbfp/image/image.html)

| FEATURES                         |  |
|----------------------------------|--|
| source                           |  |
| Seq primer: -400P from Gibco     |  |
| High quality sequence stop: 463. |  |
| location/Qualifiers              |  |
| 1. .473                          |  |

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2494906"
/clone_lib="NCI_GAP_Kid11"
/lab_host="DH10B"
/notes="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI;
plasmid DNA from the normalized library NCI_GAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500532-1502855). Subtraction by Bento Soares and M.
Fatima Ronaldo."

```

|             |       |            |        |             |          |
|-------------|-------|------------|--------|-------------|----------|
| BASE COUNT  | 110 a | 96 c       | 145 g  | 121 t       | 1 others |
| ORIGIN      |       |            |        |             |          |
| Query Match | 19.6% | Score 423; | DB 63; | Length 473; |          |

Best Local Similarity 96.08; Pred. No. 5,1e-97;  
Matches 455; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

|    |      |   |      |
|----|------|---|------|
| QY | 1677 | tgcttaagtccaaaagggggaacccaattcaacggttaggaacagtgagtcctcgagccc          | 1736 |
| Db | 471  | ttcttaaaagtcctcaaaagggggaaaacccaattcac - cgtttagaaacagtgagtcctcgagccc | 413  |
| QY | 1737 | acctcgtagacacaagaagcccgagaccagccttgagagagagccacatgacacacacagat        | 1796 |
| Db | 412  | acctcgtagacacaaanagctcgagatc - agcttgagagagagccacacactacacacacagat    | 355  |
| QY | 1797 | ggcatccttgggaagcctgaaatctatcaaccagaatctcaactcccttggccctgaac           | 1856 |
| Db | 354  | ggcatccttgggaagcctgaaatctatcaaccagaatctcaactcccttggccctgaac           | 295  |
| QY | 1857 | agggcacagaataagagaacagctcgccgacatttttgaagaagcaatgtaggaagaagag         | 1916 |
| Db | 294  | agggcacagatpaaagagacagctcgccgacatttttgaagaagcaatgtaggaagaagag         | 235  |
| QY | 1917 | cagccagccglttggagagaagatctcaagatcagcagatctcatcttctctctgccc            | 1976 |
| Db | 234  | cagccagccatpaaagagagagatctcaagatcagcagatctcatcttctctctgccc            | 175  |
| QY | 1977 | agtgatatttgctctctccagcttggggagatctctcttgaaccctaataagagccca            | 2036 |
| Db | 174  | agtgatatttgctctctccagcttggggagatctctcttgaaccctaataataagagccca         | 115  |
| QY | 2037 | cctgagatctctctctccatccctctctctctctgcccctctgctctaattgctgcagaatt        | 2096 |
| Db | 114  | ctggagatctctctctccatccctctctctctctgcccctctgctctaattgctgcagaatt        | 55   |
| QY | 2097 | gtcaatccaaacctactctctgagctcattataaataaacaagattattttcc                 | 2150 |
| Db | 54   | gtcaatccaaacctactctgagctcattataaataaacaagattattttttcc                 | 1    |

| RESULT     | 8   |
|------------|---|
| LOCUS      | W22160  |
| DEFINITION | W22160 636 bp mRNA EST 06-MAY-1996                                  |
| ACCESSION  | 6366 Human retina cDNA Tsp5091-cleaved sublibrary Homo sapiens cDNA |
| VERSION    | not directional, mRNA sequence.                                     |
| KEYWORDS   | W22160  |
| SOURCE     | W22160.1 GI:1299007   |
| ORGANISM   | EST.  |
| REFERENCE  | human.  |
| AUTHORS    | Homo sapiens  |
| TITLE      | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;       |
| JOURNAL    | Eutheria; Primates; Carnivora; Hominoidea; Homo.                    |
| COMMENT    | 1 (bases 1 to 636)  |
|            | Macke, J., Smallwood, P. and Nathans, J.                            |
|            | Adult Human Retina cDNA   |
|            | Unpublished (1996)  |
|            | On Apr 14, 1993 this sequence version replaced gi:785820.           |

Contact: Dr. Jeremy Nathans  
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics  
Johns Hopkins School of Medicine  
725 North Wolfe Street, Baltimore, MD 21205  
Tel: 410 955 4678  
Fax: 410 614 0827  
Email: jeremy-nathans@gmail.bs.jhu.edu  
Clones from this library are NOT available.  
PCR Primers  
FORWARD: CTTTGGACCAAGTCACCGCGTTAACT  
BACKWARD: GAGGTGCGTTATGATGATTTCTCCAGGGTAA  
Seq primer: GGGTAAAAAGCAAAAGATT.

```

FEATURES
SOURCE
Location/Qualifiers
1. .636
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="Human retina cDNA Tsp509i-cleaved sublibrary"
/sex="mixed (males and females)"
/tissue-type="retina"
/dev-stage="adult"

```

|            |       |       |       |       |           |
|------------|-------|-------|-------|-------|-----------|
| BASE COUNT | 141 a | 156 c | 185 g | 137 t | 17 others |
| ORIGIN     |       |       |       |       |           |

| QY  | Db     | QY  | Db    |
|-----|--------|-----|-------|
| 474 | CGGGTC | 534 | CGATG |
| 475 | CGGGTC | 535 | CGATG |
| 476 | CGGGTC | 536 | CGATG |
| 477 | CGGGTC | 537 | CGATG |
| 478 | CGGGTC | 538 | CGATG |
| 479 | CGGGTC | 539 | CGATG |
| 480 | CGGGTC | 540 | CGATG |
| 481 | CGGGTC | 541 | CGATG |
| 482 | CGGGTC | 542 | CGATG |
| 483 | CGGGTC | 543 | CGATG |
| 484 | CGGGTC | 544 | CGATG |
| 485 | CGGGTC | 545 | CGATG |
| 486 | CGGGTC | 546 | CGATG |
| 487 | CGGGTC | 547 | CGATG |
| 488 | CGGGTC | 548 | CGATG |
| 489 | CGGGTC | 549 | CGATG |
| 490 | CGGGTC | 550 | CGATG |
| 491 | CGGGTC | 551 | CGATG |
| 492 | CGGGTC | 552 | CGATG |
| 493 | CGGGTC | 553 | CGATG |
| 494 | CGGGTC | 554 | CGATG |
| 495 | CGGGTC | 555 | CGATG |
| 496 | CGGGTC | 556 | CGATG |
| 497 | CGGGTC | 557 | CGATG |
| 498 | CGGGTC | 558 | CGATG |
| 499 | CGGGTC | 559 | CGATG |
| 500 | CGGGTC | 560 | CGATG |

|    |     |   |       |
|----|-----|---|-------|
| QY | 654 | tcaatttgatgcggaatctggggcgctgcagagaaagcttgcctcccttgaaatcgga        | 72.3  |
| Db | 255 | TCAGTTGTGTGCGGAATCTGGGGCGTGCAGCAGCCCTGGGCTCCCTTGAAATCGGGA         | 31.4  |
| QY | 714 | aacaagaagaagaagtcctcatcagccctgctgagctgccagacctgcctgcagcat         | 77.3  |
| Db | 315 | AAACAAGAGAGAAGAGCTTCATCAACCTGTGTAGCGCTCCGAGGCTTGCTGCACACAT        | 37.4  |
| QY | 774 | ttccagcggccccaatccagaagcctgtgcattctatacagcaatgtg-aaacctgcgccc     | 83.22 |
| Db | 375 | TTCCAGCGGNCCTATCCGAAAGNCTGGCATCTTTATAGCCATGTGAAAACTGGCTCCC        | 43.4  |
| QY | 833 | tgctc-tgctgaagtgtagatttgcagataaggggaccagattgctgaagtcacatggcgtgcac | 89.1  |
| Db | 435 | TGCTTGCTGAGTGGGGATTGTGAATATGGGACCAAGATTGTCCAAATCATGTGNGTCAT       | 49.4  |
| QY | 892 | ttctctaactcgatccacaagaagcctgtaatatgtg                             | 92.7  |
| Db | 495 | TCCTTTAACTGTGATTCACAAAGAGAGNTGTAAATGTG                            | 53.0  |

| RESULT     | 9  |                                      |              |            |               |
|------------|--|--------------------------------------|--------------|------------|---------------|
| LOCUS      | AI628985/c   |                                      |              |            |               |
| DEFINITION | AI628985   | 411 bp                               | mRNA         | EST        | 23-APR-1999   |
| ACCESSION  | U799c2.x1  | NCI_CGAP_Kid1                        | Homo sapiens | cdna clone | IMAGE:2285282 |
| VERSION    | AI628985   |                                      |              |            |               |
| KEYWORDS   | AI628985.1   | GI:4665785                           |              |            |               |
| SOURCE     | EST.   |                                      |              |            |               |
| ORGANISM   | human.   |                                      |              |            |               |
|            | Homo sapiens   |                                      |              |            |               |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;    |                                      |              |            |               |
|            | Eutheria; Primates; Catarrhini; Homnidae; Homo.                  |                                      |              |            |               |
| REFERENCE  | 1 (bases 1 to 411)   |                                      |              |            |               |
| AUTHORS    | NCI-CGAP   | http://www.ncbi.nlm.nih.gov/ncicgap. |              |            |               |
| TITLE      | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), |                                      |              |            |               |
|            | Tumor Gene Index   |                                      |              |            |               |
| JOURNAL    | Unpublished (1997)   |                                      |              |            |               |
| COMMENT    | On Mar 10, 1998 this sequence version replaced gi:2948474.       |                                      |              |            |               |
|            | Contact: Robert Strausberg, Ph.D.                                |                                      |              |            |               |
|            | Tel: (301) 496-1550  |                                      |              |            |               |
|            | Email: Robert_Strausberg@nih.gov                                 |                                      |              |            |               |
|            | Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R |                                      |              |            |               |
|            | Emmett-Buck, M.D., Ph.D.   |                                      |              |            |               |

Seq primer: -40UP from Gibco

| FEATURES | Location/Qualifiers |
|----------|---------------------|
| SOURCE   | 1. .411             |

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone IMAGE:2285282"  
/clone\_id="NCL\_CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Plasmid DNA from the normalized library NCL\_CGAP\_Kid3 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1322376-1323911, 1456007-1456775, and 1500552-1502855). Subtraction by Bento Soares and M. Fatima Bonaldo."

|            |      |      |       |       |
|------------|------|------|-------|-------|
| BASE COUNT | 95 a | 83 c | 131 g | 102 t |
| ORIGIN     |      |      |       |       |

|                           |        |                    |           |             |
|---------------------------|--------|--------------------|-----------|-------------|
| Query Match               | 18.6%  | Score 403;         | DB 49;    | Length 411; |
| Best Local Similarity     | -98.8% | Pred. No. 5.8e-92; |           |             |
| Matches 406; Conservative | 0;     | Mismatches 5;      | Indels 0; | Gaps 0;     |

|    |      |   |      |
|----|------|---|------|
| QY | 1734 | ccacacccgfygaacaacaagaacccctccggaaccaagccttgaagagagccacatgaacaacacaca | 1793 |
| Db | 411  | CCCACTCTGTGAACACAAAGCCTTGGACCAAGCCTTTAGAGAGAGCCACACTACACACACCA        | 352  |
| QY | 1794 | gatgcacaccttggagactgaactatcatcacccaggaatctcaactccctttgagccctga        | 1853 |
| Db | 351  | GATGCATCTTGGAGACTGAATCTATACACCGAGAACTCAAACTCCCTTTGGCCCTGA             | 292  |
| QY | 1854 | accagagccacagataaaggaacagctctgggcccctttttgaagccaatgtgaggaagaag        | 1913 |
| Db | 291  | ACCAAGGCGCCAGATTAAGGAACAGCTCGGAGCCACTCTTGAAGGCCAATGTGAGAGGAAG         | 232  |
| QY | 1914 | gagagagccagccgttttgggaggaagatcctaagaatccagactatcttcctctctg            | 1973 |
| Db | 231  | GAGAGGCGAGCCGTTTGGGAGGAAGATCTCAAGGATCAACACTCAATTCCTTCCCTTGG           | 172  |
| QY | 1974 | ccceagtaattgtctctctcccaagccttggggagaccccttccttgaaccctaataagac         | 2033 |
| Db | 171  | CCCACTGAATTTGGTCTCTCTCCAGCTCTGGGGGACCTCTTCTTGAACCTTAATAAGAC           | 112  |
| QY | 2034 | ccactgagatctctctctctccatccctctctctctgcccctgtcctaattgtctgcag           | 2093 |
| Db | 111  | CCACTGGAGTCTCTCTCTCCATCCCTCTCCCTCTGCTCTAATTTGCTGCAGG                  | 52   |
| QY | 2094 | atgtgcactccaaccttactctgagctgagcttaataataaacaagattta                   | 2144 |
| Db | 51   | ATTGTCACTCCAAACCTTACTCTGAGCTCAATTAAATTAACAAGATTTA                     | 1    |

|            |  |
|------------|--|
| RESULT     | 10   |
| A1629007/c |  |
| LOCUS      | A1629007   |
| DEFINITION | A1629007 411 bp mRNA   |
|            | t7y9e02.x1 NC1_CGAP_K1d11 Homo sapiens CDNA clone IMAGE:2285306 3' |
| ACCESSION  | A1629007   |
| VERSION    | A1629007.1 GI:4665807  |
| KEYWORDS   | EST.   |
| SOURCE     | human.   |
| ORGANISM   | Homo sapiens   |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;      |

REFERENCE  
1 (bases 1 to 411)  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eutheria; Primates; Catarrhini; Homnidae; Homo.  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Mar 16, 1998 this sequence version replaced gi:2961764.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLN at:  
[www-bio.linnl.gov/dbtrp/image/image.html](http://www-bio.linnl.gov/dbtrp/image/image.html)

FEATURES  
source

Seq primer: -40UP from GIBCO.  
Location/Qualifiers  
1..411  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2285306"  
/clone\_lib="NCI CGAP\_Kid11"  
/lab\_host="DH10B"  
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with  
a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;  
Plasmid DNA from the normalized library NCI-CGAP\_Kid3 was  
prepared, and ss circles were made in vitro. Following HAP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids 1322376-1323911, 1456007-1456775, and  
1500552-1502855). Subtraction by Bento Soares and M.  
Fatima Bonaldo."

BASE COUNT 95 a 83 c 131 g 102 t  
ORIGIN

Query Match 18.6%; Score 403; DB 49; Length 411;  
Best Local Similarity 98.8%; Pred. No. 5.8e-92;  
Matches 406; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1734 cccactgtgaacacaaagcctcgaccagcttgagagagagccacatgacacacaca 1793  
|||||  
DB 411 cccacctgtgaacacaaagcctcgaccagcttgagagagagccacatgacacacaca 352  
OY 1794 gatgcatccttgagacatgaatcatcaccaggaatcacaactcccttgccctga 1853  
|||||  
DB 351 gatgcatccttgagacatgaatcatcaccaggaatcacaactcccttgccctga 292  
OY 1854 accagggccagataaggaacagctcgccacattttgaaggccaatgtgaggaag 1913  
|||||  
DB 291 accagggccagataaggaacagctcgccacattttgaaggccaatgtgaggaag 232  
OY 1914 gagcggccagcggcttgaggagaagatcgaagatccagacatctattccttctcgg 1973  
|||||  
DB 231 gagcggccagcggcttgaggagaagatcgaagatccagacatctattccttctcgg 172  
OY 1974 cccaatgaattgtctctccagcttgaggagactccttctgacccaataagacc 2033  
|||||  
DB 171 cccaatgaattgtctctccagcttgaggagactccttctgacccaataagacc 112  
OY 2034 ccaactgagactctctctccatccctctcctgacctgtcctaattgctgcagg 2093  
|||||  
DB 111 ccaactgagactctctctccatccctctcctgacctgtcctaattgctgcagg 52  
OY 2094 attgcatcacaacactactctgagctcatataataaataacagattta 2144  
|||||  
DB 51 attgcatcacaacactactctgagctcatataataaataacagattta 1

RESULT 11  
AI989344 407 bp mRNA EST 08-SEP-1999  
LOCUS w227e03.x1 NCI-CGAP\_Brn53 Homo sapiens cDNA clone IMAGE:2559292 3',  
DEFINITION mRNA sequence.  
ACCESSION AI989344  
VERSION AI989344.1 GI:5836225  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 407)  
NCI/NINDS-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute / National Institute of Neurological  
Disorders and Stroke, Brain Tumor Genome Anatomy Project  
(CGAP/BTGAAP), Tumor Gene Index  
Unpublished (1998)  
On Feb 18, 1999 this sequence version replaced gi:4296517.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Chris Moskaluk, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life  
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The  
I.M.A.G.E. Consortium DNA Sequencing by: Washington University  
Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/MLN at:  
[www-bio.linnl.gov/dbtrp/image/image.html](http://www-bio.linnl.gov/dbtrp/image/image.html)

FEATURES  
source

Seq primer: -40UP from GIBCO.  
Location/Qualifiers  
1..407  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2559292"  
/clone\_lib="NCI CGAP\_Brn53"  
/tissue\_type="three pooled meningiomas"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-Sport6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies."

BASE COUNT 97 a 81 c 127 g 102 t  
ORIGIN

Query Match 18.2%; Score 394.2; DB 63; Length 407;  
Best Local Similarity 98.0%; Pred. No. 9.9e-90;  
Matches 399; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1744 gaacacaaagcctcgaccagccttgagagagagccacatgacacacacagatgcatcc 1803  
|||||  
DB 407 gaacacaaagcctcgaccagccttgagagagagccacatgacacacacagatgcatcc 348  
OY 1804 ttggagactgaatcatatccacaggaatcgaatcccttgagcccgaaacaggcca 1863  
|||||  
DB 347 ttggagactgaatcatatccacaggaatcgaatcccttgagcccgaaacaggcca 288  
OY 1864 gataaggaacagctcgccacattttgaaggccaatgtgaggaaggaagcagccag 1923  
|||||  
DB 287 gataaggaacagctcgccacattttgaaggccaatgtgaggaaggaagcagccag 228  
OY 1924 ccglttgaggagaagatcgaaggaatcgaagactcattcccttctcgtgcccagtgat 1983  
|||||  
DB 227 ccatttgaggagaagatcgaaggaatcgaagactcattcccttctcgtgcccagtgat 168  
OY 1984 ttggtctctccagcttgaggagactccttcttaaaccttaataagccacagagagt 2043  
|||||  
DB 167 ttggtctctccagcttgaggagactccttcttgaaccccttaataagccacagagagt 108  
OY 2044 ctctctctccatccctctcctctgacctgtcctaattgtgcagatgtgcatc 2103

| Db                        | 107  | CMCTCTCTCATCCCTCTCTCTCTGCGCCCTTGCGCTTAATGTCGCCAGATGTGCATC   | 48               |
|---------------------------|--|---|------------------|
| QY                        | 2104   | caaaccttactctgagctcaatataaataaacagattatttcc                 | 2150             |
| Db                        | 47   | CAAACTTACTCTGAGCTCATTAATAATAAACAGATTATTATTC                 | 1                |
| RESULT                    | 12   |   |                  |
| LOCUS                     | AI196741/c   |   |                  |
| DEFINITION                | AI196741   | 384 bp  | mRNA             |
| ACCESSION                 | AI196741   |   | EST              |
| VERSION                   | AI196741.1   |   | 16-DEC-1999      |
| KEYWORDS                  | EST.   |   | IMAGE:2298051 3' |
| SOURCE                    | human.   |   |                  |
| ORGANISM                  | Homo sapiens   |   |                  |
| REFERENCE                 | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  |   |                  |
| AUTHORS                   | Eutheria; Primates; Catarrhini; Hominiidae; Homo.  |   |                  |
| TITLE                     | 1 (bases 1 to 384)   |   |                  |
| JOURNAL                   | NCI-CCAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .   |   |                  |
| COMMENT                   | National Cancer Institute, Cancer Genome Anatomy Project (CGAP),<br>Tumor Gene Index<br>Unpublished (1997)<br>On May 18, 1998 this sequence version replaced gi:3138695.<br>Contact: Robert Strausberg, Ph.D.<br>Tel: (301) 496-1550<br>Email: Robert.Strausberg@nih.gov<br>Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.<br>Emmert-Buck, M.D., Ph.D.<br>CDNA Library Preparation: M. Bento Soares, Ph.D.<br>DNA Sequencing by: Washington University Genome Sequencing Center<br>Clone distribution: NCI-CCAP clone distribution information can be<br>found through the I.M.A.G.E. Consortium/ILM, at:<br><a href="http://www-bio.illn.gov/bhrp/imagen/image.html">www-bio.illn.gov/bhrp/imagen/image.html</a> |   |                  |
| FEATURES                  | Insert Length: 496 Std Error: 0.00<br>Seq primer: -40UP from Glbco.  |   |                  |
| SOURCE                    | Location/Qualifiers<br>1. 384<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/clone="IMAGE:2298051"<br>/clone_lib="NCI-CCAP_Kid11"<br>/lab_host="DH10B"<br>/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with<br>a modified polylinker; Site.1: Not I; Site.2: Eco RI;<br>Plasmid DNA from the normalized library NCI-CCAP Kid3 was<br>prepared, and ss circles were made in vitro. Following HAP<br>hybridization, this DNA was used as tracer in a subtractive<br>hybridization reaction. The driver was PCR-amplified cDNAs<br>from a pool of 5,000 clones made from the same library<br>(clones 1323376-1333911, 1456007-1456775, and<br>1500552-1502855). Subtraction by Bento Soares and M.<br>Fatima Bonaldo." |   |                  |
| BASE COUNT                | 95 a 75 c 120 g 94 t   |   |                  |
| ORIGIN                    |  |   |                  |
| Query Match               | 17.2%; Score 372.8; DB 60; Length 384;   |   |                  |
| Best Local Similarity     | 98.2%; Pred. No. 2.6e-84;  |   |                  |
| Matches 377; Conservative | 0; Mismatches 7; Indels 0; Gaps 0;   |   |                  |
| QY                        | 1772   | agagggcacatgacacacacagatgacatccttggagctgacatcatcacccaggaat  | 1831             |
| Db                        | 384  | AGAGGCCACACATACACACACAGATGACATCTTGAGACCTGAATCTATCACCCAGAA   | 325              |
| QY                        | 1832   | ctcaactcccttggcctctgacacagggccagataaagaagaagctggcgcaactttt  | 1891             |
| Db                        | 334  | CTCAAACTCCCTTTTGCCCTTGAAACAGAGGCCAGATTAAGAAAGCTCGGGCACACTCT | 265              |

| QY         | 1892  | gaagggccaatgtggagggaaaggagcagccagccgtttgggaaagaatcctaagaatcga    | 1951  |
|------------|---|--|-------|
| Db         | 264   | GAAGGCCAAGCTGGAGGAAAGGAGACAGCCACTTTGGGAGAAAGATCTCAAGATCCAA       | 205   |
| QY         | 1952  | gactcatalcttccttcctctggtggccagcagtaatttgctctctccagcgtttggggactcc | 2011  |
| Db         | 204   | GACTCTCAATTCCTTTCCTCTGTGGCCAGTAGAATTTGGTCTCTCCAGCTCTGGGGAGCTCC   | 145   |
| QY         | 2012  | ttccttgaaccctaataagaagcccaactgtgagatctctctctcctccatctctctctgccc  | 2071  |
| Db         | 144   | TTTCCTTGAACCCCTAATAAGACCCCACTGGAGCTCTCTCTCCATCCCTCTCCCTCTCC      | 85    |
| QY         | 2072  | ctctctctaattgtgtgcagagattgttcactccaacttaactctcagactcaataa        | 2131  |
| Db         | 84  | CTCTCTCTAATTGCTGTGCCAGGATTTGTCACTCCAAACCTTACTCTGAGCTATTATATAA    | 25    |
| QY         | 2132  | ataaacagattatctccagcgtt  | 2155  |
| Db         | 24  | ATAACAGATTATTTTCCAGCTT   | 1     |
| RESULT     | 13  |  |       |
| LOCUS      | AI272081/c  |  |       |
| DEFINITION | g178f10.x1 NCI_CGAP_Ki43 Homo sapiens cDNA clone IMAGE:1865611 3', mRNA sequence.   |  |       |
| ACCESSION  | AI272081  |  |       |
| VERSION    | AI272081.1  |  |       |
| KEYWORDS   | GI:3891248  |  |       |
| SOURCE     | EST.  |  |       |
| ORGANISM   | Homo sapiens  |  |       |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;   |  |       |
|            | Eutheria; Primates; Catarrhini; Hominiidae; Homo.   |  |       |
| REFERENCE  | 1 (bases 1 to 385)  |  |       |
| AUTHORS    | NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .  |  |       |
| TITLE      | National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index   |  |       |
| JOURNAL    | Unpublished (1997)  |  |       |
| COMMENT    | On Jan 14, 1998 this sequence version replaced gi:1797673.  |  |       |
|            | Contact: Robert Strausberg, Ph.D.   |  |       |
|            | Tel: (301) 496-1550   |  |       |
|            | Email: Robert.Strausberg@nih.gov  |  |       |
|            | Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmerit-Buck, M.D., Ph.D.   |  |       |
|            | cDNA Library Preparation: M. Bento Soares, Ph.D.  |  |       |
|            | cDNA Library Arrayed by: Greg Lennon, Ph.D.   |  |       |
|            | DNA sequencing by: Washington University Genome Sequencing Center   |  |       |
|            | Clone distribution: NCI-CGAP clone distribution information can be  |  |       |
|            | found through the I.M.A.G.E. Consortium/ILMIL at:   |  |       |
|            | <a href="http://www-bio.11nl.gov/dbtrp/image/image.html">www-bio.11nl.gov/dbtrp/image/image.html</a>  |  |       |
| FEATURES   |   |  |       |
| source     | Insert Length: 568 Std Error: 0.00  |  |       |
|            | Seq primer: -400P from GIBCO  |  |       |
|            | High quality sequence stop: 379.  |  |       |
|            | Location/Qualifiers   |  |       |
|            | 1..385  |  |       |
|            | /organism="Homo sapiens"  |  |       |
|            | /db_xref="taxon:9606  |  |       |
|            | /clone_image="IMAGE:1865611"  |  |       |
|            | /clone_idb="NCI_CGAP_Ki43"  |  |       |
|            | /lab_host="DH10B"   |  |       |
|            | /note="Organ: kidney; Vector: pT7p3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7p3 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldi." |  |       |
| BASE COUNT | 94 a  | 72 c   | 118 g |
| ORIGIN     |   |  | 101 t |









**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:31:43 ; Search time 243.1 Seconds  
(without alignments)  
2969.166 Million cell updates/sec

Title: US-09-502-945-2  
Perfect score: 2885

Sequence: 1 ggaattcctctgtcgaagt.....aaacaaaagttaaatatt 2885

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID       | Description         |
|------------|-------|-------------|--------|----------|---------------------|
| 1          | 2885  | 100.0       | 2885   | 1 X40055 | Colon cancer assoc  |
| 2          | 84.6  | 2.9         | 32207  | 1 V73805 | KSHV LTR DNA (nucl  |
| 3          | 84.6  | 2.9         | 137507 | 1 V19941 | KSHV long unique c  |
| 4          | 79    | 2.7         | 799    | 1 V55831 | Nucleotide sequenc  |
| 5          | 79    | 2.7         | 9600   | 1 V21683 | Vector plasmid PCM  |
| 6          | 79    | 2.7         | 10596  | 1 051731 | Plasmid pcisEBON f  |
| 7          | 79    | 2.7         | 10596  | 1 T40348 | Plasmid pcisEBON f  |
| 8          | 79    | 2.7         | 10596  | 1 X15650 | Nucleotide sequenc  |
| 9          | 71.2  | 2.5         | 795    | 1 V55830 | FLCA insert stabl   |
| 10         | 68.4  | 2.4         | 2004   | 1 T85356 | Nephila clavipes s  |
| 11         | 66.4  | 2.3         | 424    | 1 V87399 | EST clone CD146. N  |
| 12         | 65    | 2.3         | 234    | 1 084832 | Spirocerbellar at   |
| 13         | 63.8  | 2.2         | 1032   | 1 023646 | Acetylpolymaine am  |
| 14         | 62    | 2.1         | 2338   | 1 014183 | N. clavipes draglin |
| 15         | 62    | 2.1         | 2338   | 1 V23249 | Nephila clavipes s  |
| 16         | 60.4  | 2.1         | 2744   | 1 098470 | Mispl.-containing p |
| 17         | 60.2  | 2.1         | 543    | 1 023092 | Antigen tc-7a gene  |
| 18         | 59.6  | 2.1         | 117213 | 1 V62176 | HSV-2 strain SB5 C  |
| 19         | 59    | 2.0         | 1505   | 1 055750 | Genomic clone G11F  |
| 20         | 58.8  | 2.0         | 1771   | 1 084834 | Spirocerbellar at   |
| 21         | 58.8  | 2.0         | 195    | 1 084831 | Spirocerbellar at   |
| 22         | 58.8  | 2.0         | 2000   | 1 N71065 | Gene encoding Plas  |
| 23         | 58.4  | 2.0         | 5661   | 1 V21518 | Rattus norvegicus   |
| 24         | 57    | 2.0         | 168    | 1 084833 | Spirocerbellar at   |
| 25         | 56.8  | 2.0         | 1995   | 1 014184 | N. clavipes draglin |
| 26         | 56.8  | 2.0         | 1995   | 1 V23250 | Nephila clavipes s  |
| 27         | 55.6  | 1.9         | 2784   | 1 T65102 | Luciferase gene un  |
| 28         | 55.4  | 1.9         | 2214   | 1 V22682 | New DNA sequence 1  |
| 29         | 55.4  | 1.9         | 3331   | 1 V22683 | New DNA sequence 1  |
| 30         | 55    | 1.9         | 916    | 1 T71743 | Human V3 loop HIV   |
| 31         | 55    | 1.9         | 1052   | 1 T27712 | Human pp32 cDNA. N  |
| 32         | 54.8  | 1.9         | 2462   | 1 V35473 | Human hSK3 coding   |
| 33         | 54.6  | 1.9         | 203    | 1 V17226 | SCA2 gene CAG repe  |
| 34         | 54.6  | 1.9         | 203    | 1 V30271 | Glutamine rich reg  |

|    |      |     |       |          |                     |
|----|------|-----|-------|----------|---------------------|
| 35 | 54.6 | 1.9 | 3076  | 1 V43674 | Receptor type tyro  |
| 36 | 53.6 | 1.9 | 12001 | 1 076213 | HSV L/ST region. H  |
| 37 | 53.2 | 1.8 | 165   | 1 V30274 | Glutamine rich reg  |
| 38 | 53.2 | 1.8 | 1005  | 1 X15318 | DNA encoding a the  |
| 39 | 52.8 | 1.8 | 2818  | 1 T10554 | Sheep mammary glan  |
| 40 | 52.6 | 1.8 | 15377 | 1 025975 | MH mutant porcine   |
| 41 | 51.8 | 1.8 | 506   | 1 084804 | Spirocerbellar at   |
| 42 | 51.8 | 1.8 | 3366  | 1 084803 | Spirocerbellar at   |
| 43 | 51.6 | 1.8 | 456   | 1 T69166 | Trypanosoma cruzi   |
| 44 | 51.4 | 1.8 | 15672 | 1 010613 | Rianodin receptor   |
| 45 | 51.2 | 1.8 | 2301  | 1 V20445 | Human c-trlk oncoge |

ALIGNMENTS

| RESULT | ID     | Sequence  | Score  | DB   | Length | Matches | Conservative | Mismatches | Indels | Gaps |
|--------|--------|---|--------|------|--------|---------|--------------|------------|--------|------|
| 1      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 2      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 3      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 4      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 5      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 6      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 7      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 8      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 9      | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 10     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 11     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 12     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 13     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 14     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 15     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 16     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 17     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 18     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 19     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 20     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 21     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 22     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 23     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 24     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 25     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 26     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 27     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 28     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 29     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 30     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 31     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 32     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 33     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 34     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 35     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 36     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 37     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 38     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 39     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 40     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 41     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 42     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 43     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 44     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |
| 45     | X40055 | ggaattcctctgtcgaagtcaagagagccacacagcgagcctcaacattccctcc | 100.0% | 2885 | 2885   | 100.0%  | 0            | 0          | 0      | 0    |

|    |      |  |      |
|----|------|--|------|
| QY | 121  | gagcgcgccccctgggacgcctctccctccacaactgccttgctgggcccctacgacag        | 180  |
| Db | 121  | gAGGgGgGcCcCcCtGgGgAcGcCtTcCtTCAAAAGtGcCtTtGcTtGgGcCtTCAcGAcAG     | 180  |
| QY | 181  | tcgagagacactccccctccgcaaaacagccctcaacccaactggaagctgac              | 240  |
| Db | 181  | TcGAGAcGAcTtCcCcCtCcGcCAAAAcCtTcTGAACCCAcTtGAAGtGcCtTAAg           | 240  |
| QY | 241  | gctaaacagaagctgtgcttaagcggagaagcagtcocctccctgcgcgcgaagga           | 300  |
| Db | 241  | GCTAAAcAGAGtGcTtGcTtAGGcGGAAGAcAGtTcCtCtCGcGcTtGcCAAGAtGgGAc       | 300  |
| QY | 301  | tgattatagcaacttaagaagagctgttgaatcacaggtccgggcttgggcgtc             | 360  |
| Db | 301  | TGtTATTAGAcCtTtTAAAGAGAcCtTtTAGATcACAGGtGcGgGcCtTtGgGgGtC          | 360  |
| QY | 361  | gtccgctgtgtaaacggccaccggcctccggcccgagctctcccaacagctcccaagac        | 420  |
| Db | 361  | GTCCGtGTgTAACGcCtACCCGcCtCCGcCtCCAcGtTtCCAAAcAcTtCCCAAGAc          | 420  |
| QY | 421  | catgcctgagaatgcttctactgtgcctcagtcgcccaacatccccaactgagatgcctccta    | 480  |
| Db | 421  | CATGcCTGAGAAAtGcTtTtACTtGcCTAcGtTcCCCAAcATCCcCAcTtGAGAtGtCtCtCA    | 480  |
| QY | 481  | gcaccgagccctccctctgtgacagctcccccacaagttcaagctctacacgtctcctc        | 540  |
| Db | 481  | GCACcGAGcCtCtCCtCTGGAAGAcTtCCcCAAcGtTtCAAcCtCTtACAGTtCtCtTTC       | 540  |
| QY | 541  | ctctgccaacatctcccttaaggtctgaaggcacaagttacactgacacaactacaacctac     | 600  |
| Db | 541  | TCTGcCCAAAcTtCtCCtTAAGGcTtGAcAGcCAcAGcTtAcTtGAcCAAcTtCAAcCtTAC     | 600  |
| QY | 601  | tgccctcccggaagctgtgcgcacagcaggaggccgagagcagccctccagtcctgcgcg       | 660  |
| Db | 601  | TGcCtCCcCGAAGcTtGtGCAGAcAGcAGGAGGcGcAGAGGcGcCtTtCAAGTcCtCtGcG      | 660  |
| QY | 661  | gcagaggtgacagctgacgcggcaagttcatagcacatcccttatctctgtgcgtgcgt        | 720  |
| Db | 661  | GCAAGGtGtGcACGcTtGAcCCcGCAAGTtCAAtGAGAcATcCtTAAAtCtTtGcCtGcTGT     | 720  |
| QY | 721  | ggcggtgcaactggagggcgagggcccccaaggagcctgcctccctgcgtgcagacatg        | 780  |
| Db | 721  | GGcGcGTGcCAcTtGAGGcGcCAcGAGAcCCcCAcGcGcATGcCtCtCGcTtGcAcAGT        | 780  |
| QY | 781  | gctcttgctgagagagcccgcgagcagagacacccctcatctgtctgtgcacctccacggca     | 840  |
| Db | 781  | GcTtGcTtGcTtGAGAGcGcCccGcGcAGAGAcACcCtCAAtTtGcTtGcCAcTtCAcGcGcCA   | 840  |
| QY | 841  | gtccccaactagtgacgggtgaacgctgtgcccacagcatlgcgacgctataggcaagctccc    | 900  |
| Db | 841  | GTCCcCAcTtAGTGAcGGGtGAAGcGTtGcGcCAcCAcATGcGAGAGcGTtAGcCAAGcTtCC    | 900  |
| QY | 901  | ggcgacatcggcctcttgccgcgaactcaagtccctacccgtcgcgcgagagttcccagggcct   | 960  |
| Db | 901  | GGcGcATtGcCtTtGAGcGcGcCAcTAcGtCtTACcCgTtGcGcGcAGAGtCCcCAAGcCtT     | 960  |
| QY | 961  | gcagcagctgtctcatgcaaacacagcaacagcagttccctgcggagaagcaagcaagca       | 1020 |
| Db | 961  | GcAGAcGcTtGcTtCTGcCAAcACAGcACcAGcAGcATTcCtTtGAGAAAGcAGAAAGcAGcAGCA | 1020 |
| QY | 1021 | gctacagcttgggcaaga tccctccaccaagacagggagcctgcccagagcagcccaaccacca  | 1080 |
| Db | 1021 | GcTtACAGcTtGGcCAAGAtTcCtCAcCAAGAcAGGcGAGcTtGCCcAGcAGcCCAcCAcCA     | 1080 |
| QY | 1081 | ccctgagagagacagagagagcctgaacggagcagcagaggtcctgctgggggaaggagc       | 1140 |
| Db | 1081 | CCcTtGAGAGAcCAAGAGAGcAGcTtGAAGcGAGcAGcAGGAGcTtTtGcTtGcGgGAGGcAGc   | 1140 |
| QY | 1141 | ccctaacctgcccggagagggctccacaagagatgagagacacacagaggaagaccttgagaga   | 1200 |
| Db | 1141 | CcTtGAcCAcTtGcCccGGAGAGGcCtTtCAcAGAGcTtGAGAcACAcAGAGAAAGcCtTtGAGAG | 1200 |
| QY | 1201 | ggagagcagagaaagagatgaggagagagagagagatgtcatccaggtttaaagacagaga      | 1260 |

|    |      |  |      |
|----|------|--|------|
| Db | 1201 | GGAGAGACGAGAAAGAGATGGGAGAGAGAGAGATTCATCCAGTTAAAGACGAGGA            | 1260 |
| Qy | 1261 | ggcgagagtggtgtctcgaagagagggccgcgacttggagagagctgtgtctgtatatacaaaa   | 1320 |
| Db | 1261 | GGCGGAGAGTGCTCTGAGGAGGGGGCCGCACTTTGGAGAGAGCCTGGTGTGGATACAAAA       | 1320 |
| Qy | 1321 | acgtlctcagatgtgccacacgcgtlgaaccttctgcaagtgtaaccaagcgccctcaagct     | 1380 |
| Db | 1321 | ACTGTTCTCAGATGCCCAACCGCTGCACCTTTGGCAGGTGTACCAAGCGCCCTCACACT        | 1380 |
| Qy | 1381 | ggccacgtgtcccccaccaagccttggcgctgataccaatctcccttgtgtcccttggggg      | 1440 |
| Db | 1381 | ggccactgtcccccacccacccctctggagcggtraccCAATCTCCCTGTGCCCCCTGGGGG     | 1440 |
| Qy | 1441 | catgaagaccccccaacaaacccgtlcaagcactcttcaacacaagtggtgtctaga          | 1500 |
| Db | 1441 | CATGAAGACCCCCCAACCAACCCGTCAGACACTCTTACACCAAACTGTGTGTACGA           | 1500 |
| Qy | 1501 | caactlcatgtctaaagcaacagtgcatgtgtcgggaaacacacacgltgcacccctgacatgc   | 1560 |
| Db | 1501 | CACGTTCTGTCTAAAGCACCAAGTCATGTGTGGGAAACACACGCTGCACCTGTGAGCATGC      | 1560 |
| Qy | 1561 | ttgcccgaatccagagcatctgtgtcccgctctgacagagaaagcctgtcttaagaagtgcga    | 1620 |
| Db | 1561 | TGGCGGATCCAGAGCACTGTGTGTCCGGCTGCAGGAGACAGGCTGTCTTACCAAGTGCGA       | 1620 |
| Qy | 1621 | ggcgatccgaggtgtgcacaagccacgctltagatgtcagaagtgcatctgtgatacca        | 1680 |
| Db | 1621 | GCGGATCCGAGGTGTGCAAACCCAGCTAGATGATGCCAGACAGTGCACCTGTGAATACA        | 1680 |
| Qy | 1681 | cacctgtctctatgtggaaccgactccctlcaaccgycagaaagttagacagacaagaattgtct  | 1740 |
| Db | 1681 | CACCTGTCTATGTGGACACAGTCCCTCTCAACCGCGCAGAGCTAGACACAAGAAGTTGCT       | 1740 |
| Qy | 1741 | cggtlccatcaagccaagaagtgtatgtctgtcgtccttgttggggcatccggggtggagag     | 1800 |
| Db | 1741 | CGGTCCCATACGCCAGAAAGTGTATGTCTGTCTTGTGGGGCATCCGGGAGTGGAGAG          | 1800 |
| Qy | 1801 | tgacaacgttggagatagaatgcacccctccagttgtgtgtgcacatgagagtggtcgtcct     | 1860 |
| Db | 1801 | TGACACCGTGTGGAAATAGATGCACACTCTTCAGTGTGTGCCCATAGGCAGTGGCTGCT        | 1860 |
| Qy | 1861 | gctgtagcttgccttcaaggttgctctgacagagagtcctcaagaaatggatttgcacatccg    | 1920 |
| Db | 1861 | GCTGAGCTGTGCTTCAAGGTGTGCTGAGAGAGTGCAGAAATGATTTGCCCATCATACCG        | 1920 |
| Qy | 1921 | ggccccagagacacagcgccgaggaatccacagccatgggaattctgtcttcttcaactcgt     | 1980 |
| Db | 1921 | GCCCCAAGACACACCCCGAGGAATCCAGCCAGGGAATTCCTCTTTCATCACTGCT            | 1980 |
| Qy | 1981 | agccaaccgcaaaactctcagacagaagaatttgaacgttggcgcaagrtcccaatcgttga     | 2040 |
| Db | 1981 | AGCCATCAACCCAAAATCTCTTACAGCAGAAATTGAACGTGGCCAAAGTCTCTCATGTGGA      | 2040 |
| Qy | 2041 | ctggaacatltaccatlgtgcaatgtggcacccacagcagcgcttctcaaatgaacccctcgtgtc | 2100 |
| Db | 2041 | CTGGGACATTCACATGGCAATGGCAACCCAGCGCTTCTCAATGAACCCCTGTGTCT           | 2100 |
| Qy | 2101 | ctcaatctctctgcatcgtctatgacaacgaggaacttcttcagagctcttgggctcttga      | 2160 |
| Db | 2101 | CTACATCTCTCTGTGATGCTATGATCAACAGGGAATCTTCTTCCAGGCTCTGGGGGCTCTGA     | 2160 |
| Qy | 2161 | agaggttgggtggagagaccagcgltgggtgtacaaatfgaaacgttggaaatggaaagagtggt  | 2220 |
| Db | 2161 | AAAGTGTGTGTGAGACCGAGCGTGGGGTGTCAATGTGAACGTGGCATGGACAGGAGTGT        | 2220 |
| Qy | 2221 | gaaccccccaatlgagagcgttggagtaccttcaagccttcaagagacgttgggtgatgccat    | 2280 |
| Db | 2221 | GAACCCCCATTGGAAACGTGGAGTACTTACAGCTTCAAGACACGTGTGATGCCAT            | 2280 |
| Qy | 2281 | ttgccacagattctaccatgatgtgtcctagtctccgcgggtttgatatgcttgaag          | 2340 |



|    |   |  |
|----|---|--|
| FT | CDS   | 1142. .2794                                  |
| FT | /**tag= a   | /product= complement-binding protein         |
| FT | 8699. .11236  |  |
| FT | /**tag= b   | /product= glycoprotein B                     |
| FT | complement(17261. .17875)   |  |
| FT | /**tag= c   | /product= interleukin 6                      |
| FT | complement(21548. .21832)   |  |
| FT | /**tag= d   | /product= macrophage inflammatory protein II |
| FT | complement(27137. .27424)   |  |
| FT | /**tag= e   | /product= interferon regulatory factor 1     |
| FT | 28661. .29741   |  |
| FT | /**tag= f   | /product= protein T1.1                       |
| FT | complement(58976. .60175)   |  |
| FT | /**tag= g   | /product= glycoprotein M                     |
| FT | complement(69412. .69915)   |  |
| FT | /**tag= h   | /product= glycoprotein L                     |
| FT | complement(88410. .88910)   |  |
| FT | /**tag= i   | /product= interferon regulatory factor 2     |
| FT | 89600. .90541   |  |
| FT | /**tag= j   | /product= interferon regulatory factor 3     |
| FT | 90173. .90643   |  |
| FT | /**tag= k   | /product= glycoprotein X                     |
| FT | complement(93636. .94127)   |  |
| FT | /**tag= l   | /product= interferon regulatory factor 4     |
| FT | complement(111931. .112443)   |  |
| FT | /**tag= m   | /product= capsid protein IV                  |
| FT | complement(123808. .127296)   |  |
| FT | /**tag= n   | /product= immediate early protein            |
| PD | M09804576-A1.   |  |
| PD | 05-FEB-1998.  |  |
| PF | 22-JUL-1997; U13346.  |  |
| PR | 29-NOV-1996; US-757669.   |  |
| PR | 25-JUL-1996; US-686243.   |  |
| PR | 25-JUL-1996; US-686349.   |  |
| PR | 25-JUL-1996; US-686350.   |  |
| PR | 25-JUL-1996; US-687253.   |  |
| PR | 25-JUL-1996; US-688814.   |  |
| PR | 05-SEP-1996; US-708678.   |  |
| PR | 10-OCT-1996; US-728323.   |  |
| PR | 13-NOV-1996; US-747887.   |  |
| PR | 13-NOV-1996; US-748640.   |  |
| PA | (UTCO ) UNIV COLUMBIA NEW YORK.   |  |
| PI | Bolzhenky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;                      |  |
| PI | WPI; 98-130615/12.  |  |
| PT | New nucleic acid encoding Kaposi's sarcoma associated herpes virus          |  |
| PT | protein - useful for e.g. detecting levels of HHV8 in, and                  |  |
| PT | preparation of vaccines for treatment of, HIV patients                      |  |
| PS | Example 2; Page 135-203; 230pp; English.                                    |  |
| CC | This sequence represents the long unique region and terminal repeat of      |  |
| CC | the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known     |  |
| CC | as human herpes virus 8 (HHV8). This sequence contains the DNAs of the      |  |
| CC | invention which encode KSHV polypeptides selected from: (a) viral           |  |
| CC | microphage inflammatory protein-(binding) protein; glycoproteins B, M or L; |  |
| CC | (c) viral INF 1; (d) complement-binding protein; glycoproteins B, M or L;   |  |
| CC | (d) capsid protein IV encoded by ORF6; and (e) immediate early protein      |  |
| CC | encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded    |  |
| CC | by it, and antibodies (Ab) specific for the proteins are useful for         |  |
| CC | detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body     |  |
| CC | fluids or tissue samples. HHV8 infections can be treated with antisease     |  |

or triplex forming molecules or agents that bind specifically to the protein. Ab may be used for prophylaxis or treatment of HHV8 infection, while the protein can be used in protective vaccines. Ab may also be used to differentiate between lymphomas, and HHV8 may be implicated in many other lymphoproliferative diseases such as lymphomas, leukaemia, splenomegaly and mycosis fungoides. Cells and animals containing the nucleic acid are useful for drug screening. HHV8-derived peptides can be used as targets for antiviral drugs, e.g. dihydropyridine reductase gene can be inhibited with methotrexate. These can also be used to determine the immune status of a patient infected with HIV. HHV8 derived protein viral MIP III may be used as an anti-inflammatory agent for, e.g. treating rheumatoid arthritis. This sequence is stated as containing 81 open reading frames.

Sequence 137507 BP, 32579 A, 37795 C, 35758 G, 31375 T;

|                       |              |                  |                 |                |
|-----------------------|--------------|------------------|-----------------|----------------|
| Query Match           | 2.9%;        | Score 84.6;      | DB 1;           | Length 137507; |
| Best Local Similarity | 48.8%;       | Pred. No. 7e-09; |                 |                |
| Matches 293;          | Conservative | 0;               | Mismatches 299; | Indels 9;      |
|                       |              |                  |                 | Gaps 2         |

|        |  |  |        |
|--------|--|--|--------|
| OY     | 770  | ctgaagatggtcgtttgtcttgagcaaggccggacagaagacaaccctattgtctgtcca         | 829    |
| Db     | 125664   | CAGCAGCAGGAGCCACACACACACAGAGGCACACAGCAAGAAGAGCTCACAGCAGCGGGAACTCA    | 125635 |
| OY     | 830  | cttcaagcgagcttcctccactagttaacggtgtgaacgtgtgtgcaccaacgaatctcgtgcagcta | 889    |
| Db     | 125634   | CAGCAGCGGGAGGCCACAGCAGCGGGAGCCACAGCAGAGGGGGAGCCACAGCAGCGGGAACTCA     | 125575 |
| OY     | 880  | ggcaagctcccgcgagcatcggcccccttagccgcaactagtcttcaaccgctgtgcgcag---     | 947    |
| Db     | 125574   | CAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAACTCA        | 125515 |
| OY     | 947  | agtcctccagggccctgcagcagctgtgatcatgaaacaacagcaccaagcttctcttgagaag     | 1006   |
| Db     | 125514   | CAGCAGCAGGATGTAGACAGCAGCAGAGGATGTAGCAGCAGCAGCATGAGTCAGCAGCAGCAGATGAG | 125455 |
| OY     | 1007   | cagaagcagcagcagctaacagcttggccaagatctctcaaccaagacagggagctgtgccag      | 1066   |
| Db     | 125454   | CAGCAGCAGGATGTAGACAGCAGCAGAGGATGTAGCAGCAGCAGATGTAGCAGCAGCAGATGAG     | 125395 |
| OY     | 1067   | cagcccaaccaaccacctgaggaagacaagaggaagcttgaacgagcagcaggggtctcttg       | 1126   |
| Db     | 125394   | CAGCAGCAGGATGTAGACAGCAGCAGAGGATGTAGCAGCAGCAGATGTAGCAGCAGCAGAGATGAG   | 125335 |
| OY     | 1127   | ctgggggagggagcccttgaccatgtccctccgggaggggtctcaagaagattgaagacaacag     | 1186   |
| Db     | 125334   | CAGCAGCAGGATGTAGC-----A CAGCAGAGGATGTAGCAGCAGCAGCAGGATGTAGCAGCAGCAG  | 125281 |
| OY     | 1187   | gaagaccttgagagagagcagcaggaagaaagatatgtgggagagagagagatgtcatccag       | 1246   |
| Db     | 125280   | GATGAGCAGCAGCAGGATGTAGACACACAGCAGCAGAGGATGTAGCAGCAGCAGCAGCAGCAG      | 125221 |
| OY     | 1247   | gttaagacagcagggagcgcgagatggtgtcctgaggaagggcccgacttgtggagggccggt      | 1306   |
| Db     | 125220   | GATGAGCAGCAGCAGGATGTAGACACAGCAGCAGGATGTAGCAGCAGCAGCAGCAGATGTAGCAG    | 125161 |
| OY     | 1307   | gctgatacaaaaacatgttctcagatgtgcccaacgcgtgaaccttgcaggtgtacca           | 1366   |
| Db     | 125160   | GAGCAGCAGGATGTAGCAGCAGCAGCAGAGATGTAGCAGCAGCAGCAGCAGATGTAGCAGCAGCAG   | 125101 |
| OY     | 1367   | g  | 1367   |
| Db     | 125100   | G  | 125100 |
| RESULT | 4  |  |        |
| ID     | V55831   |  |        |
| AC     | V55831 standard; DNA; 799 BP.  |  |        |
| DT     | 18-NOV-1998 (first entry)  |  |        |
| DE     | Nucleotide sequence of the stabilising sequence-encoding insert.   |  |        |
| KW     | Fusion protein; stabilising polypeptide; proteolytic degradation; resistance; half-life; autoimmune disease; inflammation; nitro drug; |  |        |







PS Example 1; Fig 6; 39pp; English.  
CC This sequence represents the episomal CMV driven expression plasmid  
CC pcisEBON which was used in the expression of variant human hepatocyte  
CC growth factor (HGF). HGF is isolated from human serum and is a  
CC disulphide linked heterodimer derived by proteolytic cleavage of the  
CC pro-hormone between residues 494 and 495. This generates a molecule  
CC composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and  
CC a beta subunit of 234 amino acids (mol. wt. 34 kD). The alpha subunit  
CC subunits are encoded by a single open reading frame. The alpha subunit  
CC contains four kringle domains based on their homology to kringle-like  
CC domains in other proteins, e.g. prothrombin, plasminogen. The beta  
CC subunit shows high homology to the catalytic domain of serine proteases.  
CC However two of the three residues which form the catalytic triad of  
CC serine proteases are not conserved in HGF. Therefore, the precise  
CC function of the beta chain remains unknown. The invention includes HGF  
CC variants which retain HGF receptor binding activity without having the  
CC biological activity of wild-type HGF. They can be used for the treatment  
CC of pathological conditions associated with the activation of a HGF  
CC receptor such as malignancies associated with chronic HGF receptor  
CC activation. The pcisEBON plasmid comprising the HGF coding sequence  
CC may be used for manipulation of the HGF coding sequence and expression  
CC of the variant HGF's of the invention.  
SQ Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;

Query Match 2.7%; Score 79; DB 1; Length 10596;  
Best Local Similarity 51.3%; Pred. No. 5.3e-08;

Matches 184; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

OY 943 gcagagtcctccagccctcagcagctgtcatgtcaacacacagcaccagcttcttga 1002  
DB 2369 GCAGAGGGGCGAGAGGGGCGAGAGGAGGAGGCGAGAGAGAGAGGCGGCGA 2428  
OY 1003 gaagcagaagcagcagcagctcagctgtggaagatctccaaagacagggagctgc 1062  
DB 2429 GCGGCGAGGAGAGAGAGGAGGCGAGGAGGCGAGAGAGAGAGAGGCGGCA 2488  
OY 1063 caggcagccaccaccacccttgagagacagagagagagctcagagcagcagaggt 1122  
DB 2489 GGAGCAGAGAGAGAGGCGAGAGGCGAGAGGCGGCGAGAGGCGGCGAGAGG 2548  
OY 1123 ctgtctggtggagggagccctgaccatgcccgggaggggtccacagagatgtagagcac 1182  
DB 2549 GCAGAGCAGAGAGAGGCGAGAGGCGGCGAGAGGCGGCGAGAGAGAGCA 2608  
OY 1183 acaggaagaccttggagagagagcaggaagaagatgtggagagagagattgcat 1242  
DB 2609 GGAGGGGCGAGAGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGGCGCA 2668  
OY 1243 ccaggttaagagcagggagggcgagagtggtgtcagagagagggcccgacttggagagc 1301  
DB 2669 GGAGCAGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGC 2727

RESULT 8  
X15650  
ID X15650 standard; DNA: 10596 BP.  
AC X15650;  
DT 10-MAY-1999 (first entry)  
DE Nucleotide sequence of plasmid pcis.EBON.  
KW Plasmid pcis.EBON; hepatocyte growth factor; HGF; variant;  
OS HGF receptor activation; ss.  
OS Synthetic.  
PN US5879910-A.  
PD 09-MAR-1999.  
PF 09-FEB-1994; 194087.  
PR 09-FEB-1994; US-194087.  
PA (GENE) GENENTECH INC.  
PI Godowski FJ, Loker NA, Mark MR:  
DR WPI; 99-203949/17.  
PT New hepatocyte growth factor variants - useful for studying  
PT structure-function relationships in the wild type molecule and for  
PT treating conditions associated with chronic hepatocyte growth factor

PT receptor activation  
PS Example 1; Fig 6A-F; 40pp; English.  
CC The present sequence represents the nucleotide sequence of pcis.EBON,  
CC which is used in the course of the invention. The specification  
CC describes a hepatocyte growth factor (HGF) variant (HGFV) comprising an  
CC amino acid (aa) alteration at or adjacent to position 692 of the  
CC wild-type human HGF (huHGF) aa sequence. HGFV may be used in studies to  
CC identify the structure-activity relationships of HGF in order to identify  
CC the functionally important domains in the aa sequence. It may also be  
CC used to identify aa residues which are responsible for the interaction of  
CC HGF with its receptor, and those aa that are responsible for the  
CC biological activity of it. Variants of HGFV which have enhanced receptor  
CC binding affinity (compared to wild-type huHGF) and are more biologically  
CC active than wild type huHGF, may be used as huHGF agonists. Conversely,  
CC variants of HGFV which have enhanced receptor binding affinity (compared  
CC to wild-type huHGF) but which are biologically inactive may be used as  
CC huHGF antagonists, and may be used to block the binding of wild-type  
CC huHGF to its receptor. This permits the treatment of pathogenic  
CC conditions associated with the activation of an HGF receptor, such as  
CC malignancies associated with chronic HGF receptor activation.  
CC HGFV-immunoglobulin chimeras may be produced (by standard methods) and  
CC used in protein A purification, immunohistochemistry, and  
CC immunoprecipitation techniques (in place of anti-HGF antibodies) or in  
CC screening studies to identify inhibitors of HGF-HGF interactions.  
SQ Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;

Query Match 2.7%; Score 79; DB 1; Length 10596;  
Best Local Similarity 51.3%; Pred. No. 5.3e-08;

Matches 184; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

OY 943 gcagagtcctccagccctcagcagctgtcatgtcaacacacagcaccagcttcttga 1002  
DB 2369 GCAGAGGGGCGAGAGGGGCGAGAGGAGGAGGCGAGAGAGAGAGGCGGCGA 2428  
OY 1003 gaagcagaagcagcagcagctcagctgtggaagatctccaaagacagggagctgc 1062  
DB 2429 GCGGCGAGGAGAGAGAGGCGAGAGGCGGCGAGAGGCGGCGAGAGAGGCGGCA 2488  
OY 1063 caggcagccaccaccacccttgagagacagagagagagctcagagcagcagaggt 1122  
DB 2489 GGAGCAGAGAGAGAGGCGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGG 2548  
OY 1123 ctgtctggtggagggagccctgaccatgcccgggaggggtccacagagatgtagagcac 1182  
DB 2549 GCAGAGCAGAGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGAGAGCA 2608  
OY 1183 acaggaagaccttggagagagagcaggaagaagatgtggagagagagattgcat 1242  
DB 2609 GGAGGGGCGAGAGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGGCGCA 2668  
OY 1243 ccaggttaagagcagggagggcgagagtggtgtcagagagagggcccgacttggagagc 1301  
DB 2669 GGAGCAGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGGCGGCGAGAGC 2727

RESULT 9  
V55830  
ID V55830 standard; DNA: 795 BP.  
AC V55830;  
DT 18-NOV-1998 (first entry)  
DE FLGA Insert stabilising polypeptide encoding DNA.  
KW Fusion protein; stabilising polypeptide; proteolytic degradation;  
KW resistance; half-life; autoimmune disease; inflammation; nitro drug;  
KW Ikappab regulator protein; inflammatory bowel disease; in vivo imaging;  
KW nitroreductase protein; enzyme therapy; prodnug therapy; protease;  
KW cancer; pathological condition; ss.  
OS Epstein-Barr virus.  
FH Location/Qualifiers  
FT Key  
FT CDS  
FT 1..788  
FT /tag= a  
FT /product= "stabilising polypeptide"  
PN W09822577-A1.







OY 1295 gagag 1300  
||  
DB 1004 GACAG 1009

Search completed: May 20, 2000, 00:36:50  
Job time: 11862 sec

**THIS PAGE BLANK (USPTO)**







TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 2.7%; Score 58.8; DB 3; Length 32207;  
Best Local Similarity 52.4%; Pred. No. 6.8e-06;  
Matches 129; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 978 gagagcgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagctgcagc 1037  
DB 19813 GCAGCGACGACGAGGATGACGACGACGACGATGACGACGACGACGACGATGACGACG 19754  
QY 1038 gcagagcgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagcag 1097  
DB 19753 GCAGCGACGAGTGAACAGAGGACGACGAGGACGACGACGACGACGACGACGACG 19694  
QY 1098 gcagagcgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagcag 1157  
DB 19693 GTTAGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 19634  
QY 1158 gagatcccggaagagatgagacagattgtgagagcgaggaagaaatttaagaagcag 1217  
DB 19633 GCGAGTTAGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 19574

QY 1218 ggaaga 1223  
DB 19573 GCAGGA 19568

RESULT 4  
US-09-130-114-1/c  
Sequence 1, Application US/09130114  
Patent No. 5976807  
GENERAL INFORMATION:  
APPLICANT: Horlick, Robert A.  
APPLICANT: Robbins, Alan K.  
TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes  
TITLE OF INVENTION: From Multiple Transfected Episomes  
FILE REFERENCE: 0867/1D903US1  
CURRENT FILING DATE: 1998-08-06  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 5452  
TYPE: DNA  
ORGANISM: YEBNA  
US-09-130-114-1

Query Match 2.6%; Score 57.2; DB 4; Length 5452;  
Best Local Similarity 50.7%; Pred. No. 7e-06;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagagctgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagctgcag 1035  
DB 2138 CGGAGAGGCGCAGAGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2079  
QY 1036 cggagagcgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagcag 1095  
DB 2078 CAGGAGGCGGCGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2019  
QY 1096 cagcagagatgagcgcgcaagaggaanaattgtcccaagagcgagcagaggaanaat 1155

DB 2018 GGCAGAGAGGCGGCGCAGAGGCGGCGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1959  
QY 1156 gagagataccggaagagatggaacaagattgagagggaggaagaaatttaagaagcaa 1215  
DB 1958 GAGGGGCGACAGACAGAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1899  
QY 1216 tgggaagaagactggggtcacaagagacag 1245  
DB 1898 CAGGACGACGAGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1869

RESULT 5

US-07-884-811-15  
Sequence 15, Application US/07884811  
Patent No. 5316921  
GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/884, 811  
FILING DATE: 19920518  
CLASSIFICATION: 330  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dregler, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 755.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/952-9881  
TELEFAX: 415/371-7168  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-07-884-811-15

Query Match 2.6%; Score 57.2; DB 1; Length 10596;  
Best Local Similarity 50.7%; Pred. No. 1e-05;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagagctgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagctgcag 1035  
DB 2247 CGGAGAGGCGCAGAGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2306  
QY 1036 cggagagcgttcacatgacagaccgagcggtgtgagagcgccgagcagcgtgagcag 1095  
DB 2307 CAGAGAGGCGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1095  
QY 1096 cagcagagatgagcgcgcaagaggaanaattgtcccaagagcgagcagaggaanaat 1155  
DB 2367 GCGCAGAGGCGGCGCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2306  
QY 1156 gagagatcccggaagagatggaacaagattgttaagagaggaagaaatttaagaag 1155

Db 2427 GAGGGGACAGAGCAGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGAGGGG 24186  
Qy 1216 tgggaagaagacttgggtctcaaggaacag 1245  
Db 2487 CAGGAGCAGAGAGGAGGGGCGAGAGGGGCGAG 2516

RESULT 6  
US-07-885-971-15  
; Sequence 15, Application US/07885971  
; Patent No. 5328837  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.25 inch, 360 Kb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: palin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/885,971  
; FILING DATE: 19920518  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dreger, Ginger R.  
; REGISTRATION NUMBER: 33,055  
; REFERENCE/DOCKET NUMBER: 779  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-3216  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 bases  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-07-885-971-15

Query Match 2.6%; Score 57.2; DB 1; Length 10596;  
Best Local Similarity 50.7%; Pred. No. 1e-05;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
Qy 976 cggagagctgttcaatgacagaccgagagcgctgtgcagagcgcgagcgtgtgagctgcag 1035  
Db 2247 CGGGAGGGGCGAGAGCAGAGAGGGGCGAGAGCGAGAGAGGGGCGAGAGCAGAGAGGAGG 2306  
Qy 1036 cggagagagcttctcatgacagaagcgctgtgcagatgagagtcacaagaatctccagag 1095  
Db 2307 CAGGAGGGGCGAGAGGGGCGAGAGCGAGAGCGAGAGGGGCGAGAGCAGAGCAGAGCAG 2366  
Qy 1096 cagcagagatgtagcgagcaagaagaaatgtgccagaagcgagcagaggaat 1155  
Db 2367 GGGCAGAGGGGCGAGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCAGAGAGGGGCGAG 2426  
Qy 1156 gagagataccggaagagagatgtagagaggaagaagtttaagaagcaa 1215  
Db 2427 GAGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGAGGG 2486  
Qy 1216 tgggaagaagacttgggtctcaaggaacag 1245

Db 2487 CAGGAGCAGAGAGGAGGGGCGAGAGGGGCGAG 2516

RESULT 7  
US-08-087-783A-15  
; Sequence 15, Application US/08087783A  
; Patent No. 5547856  
; GENERAL INFORMATION:  
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genentech, Inc.  
; STREET: 460 Point San Bruno Blvd  
; CITY: South San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94080  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Winpatin (Genentech)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/087,783A  
; FILING DATE: 13-Jul-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/884811  
; FILING DATE: 18-MAY-92  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/885971  
; FILING DATE: 18-MAY-92  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Marschang, Diane L.  
; REGISTRATION NUMBER: 35,600  
; REFERENCE/DOCKET NUMBER: P0755779P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415/225-5416  
; TELEFAX: 415/952-9881  
; TELEX: 910/371-7168  
; INFORMATION FOR SEQ ID NO: 15:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 10596 base pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Single  
; TOPOLOGY: linear  
; US-08-087-783A-15

Query Match 2.6%; Score 57.2; DB 1; Length 10596;  
Best Local Similarity 50.7%; Pred. No. 1e-05;  
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;  
Qy 976 cggagagctgttcaatgacagaccgagagcgctgtgcagagcgcgagcgtgtgagctgcag 1035  
Db 2247 CGGGAGGGGCGAGAGCAGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCAGAGAGGAGG 2306  
Qy 1036 cggagagagcttctcatgacagaagcgctgtgcagatgtagagtcacaagaatctccagag 1095  
Db 2307 CAGGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCAGAGCAG 2366  
Qy 1096 cagcagagatgtagcgagcaagaagaaatgtgccagaagcgagcagaggaat 1155  
Db 2367 GGGCAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGCGAGAGGGGCGAG 2426  
Qy 1156 gagagataccggaagagagatgtagagaggaagaagtttaagaagcaa 1215  
Db 2427 GAGGGCGAGAGCAGAGAGAGGGGCGAGAGGGGCGAGAGGGGCGAGAGCAGAGAGGAGGG 2486  
Qy 1216 tgggaagaagacttgggtctcaaggaacag 1245  
Db 2487 CAGGAGCAGAGAGGAGGGGCGAGAGGGGCGAG 2516

```

RESULT      8
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-088B-15

```

```

Query Match      2.6%; Score 57.2; DB 1; Length 10596;
Best Local Similarity 50.7%; Pred. No. 1e-05;
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagctgttcattgacagaccgagcgctgacagcgcgagcgctgagcgag 1035
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2247 CGGAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2306
QY 1036 cggagagcttctcattgacagacgagcgctgagcgagcgagcgagcgag 1095
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2307 CAGGAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2366
QY 1096 cagcagagagatgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1155
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2367 GGGCAGGAGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2426
QY 1156 gagagataccgagagagatgagcgagcgagcgagcgagcgagcgagcgagcgag 1215
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2427 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2486
QY 1216 tgggagagagactgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2487 CAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2516

```

RESULT 9  
US-08-194-087-15

```

; Sequence 15, Application US/08194087
; Patent No. 5879910
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dieger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-087-15

```

```

Query Match      2.6%; Score 57.2; DB 3; Length 10596;
Best Local Similarity 50.7%; Pred. No. 1e-05;
Matches 137; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 976 cggagctgttcattgacagaccgagcgctgacagcgcgagcgctgagcgag 1035
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2247 CGGAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2306
QY 1036 cggagagcttctcattgacagacgagcgctgagcgagcgagcgagcgagcgag 1095
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2307 CAGGAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2366
QY 1096 cagcagagagatgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1155
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2367 GGGCAGGAGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2426
QY 1156 gagagataccgagagagatgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1215
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2427 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2486
QY 1216 tgggagagagactgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2487 CAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2516

```

RESULT 10  
PCT-US93-04648-15  
; Sequence 15, Application PC/TUS9304648  
; GENERAL INFORMATION:  
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie  
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:36:50 ; Search time 243.1 seconds

(without alignments)  
1335.867 Million cell updates/sec

Title: US-09-502-945-3

Sequence: 1 ggcctgtgaatgactgcga.....ggaagaagaagaagaanaa 1298

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query Match | Length | DB ID       | Description        |
|------------|-------|-------------|--------|-------------|--------------------|
| 1          | 1288  | 99.2        | 1298   | 1 X40056    | Colon cancer assoc |
| 2          | 191.2 | 14.7        | 6644   | 1 X33181    | Base sequence of t |
| 3          | 191.2 | 14.7        | 7372   | 1 X33182    | Base sequence of t |
| 4          | 191.2 | 14.7        | 7797   | 1 X33180    | Compo virus bsr f  |
| 5          | 191.2 | 14.7        | 7996   | 1 X33184    | Base sequence of t |
| 6          | 148.2 | 11.4        | 1686   | 1 087587    | DNA encoding Leuco |
| 7          | 145   | 11.2        | 4000   | 1 T91902    | Mannose-1-phosphat |
| 8          | 142.2 | 11.0        | 3399   | 1 T05868    | Chicken leucocytos |
| 9          | 125.6 | 9.7         | 11901  | 1 X02998    | Human IL-1ra BAC c |
| 10         | 125.2 | 9.6         | 2056   | 1 V62142    | HSV-2 strain SBS c |
| 11         | 124   | 9.6         | 11084  | 1 X22302    | Human IL-1ra BAC c |
| 12         | 118   | 9.1         | 6755   | 1 V21511    | Staphylococcal bac |
| 13         | 117.6 | 9.1         | 8952   | 1 V62130    | HSV-2 strain SBS c |
| 14         | 116.2 | 9.0         | 44453  | 1 X23519    | Human kidney amino |
| 15         | 114.6 | 8.8         | 29392  | 1 V15422    | Mouse poly Ig rece |
| 16         | 112.6 | 8.7         | 700    | 1 V88907    | EST clone 1A200. N |
| 17         | 110   | 8.5         | 114955 | 1 X53491    | Human adenosine A1 |
| 18         | 109.4 | 8.4         | 10620  | 1 X02996    | Human IL-1ra BAC c |
| 19         | 109.4 | 8.3         | 14690  | 1 X22303    | Human IL-1ra BAC c |
| 20         | 107.4 | 8.3         | 980    | 1 V06088    | Viral infection ge |
| 21         | 103.8 | 8.0         | 53526  | 1 T94101    | Human PKD1 gene. H |
| 22         | 103.8 | 8.0         | 53577  | 1 T18551    | Human polygenic k  |
| 23         | 103.8 | 8.0         | 53577  | 1 T94108    | Human PKD1 locus b |
| 24         | 99.8  | 7.7         | 12700  | 1 V62133    | HSV-2 strain SBS c |
| 25         | 98.8  | 7.6         | 5741   | 1 V62150    | HSV-2 strain SBS c |
| 26         | 98.4  | 7.6         | 4590   | 1 M60472    | Sequence encoding  |
| 27         | 98.2  | 7.6         | 9789   | 1 T41852    | CDNA encoding plas |
| 28         | 96.4  | 7.4         | 3653   | 1 X20280    | Borrelia burgdorfe |
| 29         | 95.4  | 7.3         | 18124  | 1 T21882    | Plasmodium var-7 g |
| 30         | 95    | 7.3         | 2784   | 1 T65102    | Luciferase gene u  |
| 31         | 93.6  | 7.2         | 2440   | 1 T76782    | Staphylococcus aur |
| 32         | 91.6  | 7.1         | 2447   | 1 V54587    | Human secretory pr |
| 33         | 91.6  | 7.1         | 35515  | 1 X20252    | Borrelia burgdorfe |
| 34         | 89.8  | 6.9         | 110000 | 1 V21209_06 | Continuation (7 of |

|    |      |     |       |          |                    |
|----|------|-----|-------|----------|--------------------|
| 35 | 88.8 | 6.8 | 202   | 1 V41451 | Nucleotide sequenc |
| 36 | 88   | 6.8 | 1559  | 1 T91855 | DUB-1 enhancer/pro |
| 37 | 87.4 | 6.7 | 26338 | 1 V62134 | HSV-2 strain SBS c |
| 38 | 87   | 6.7 | 2487  | 1 X20298 | Borrelia burgdorfe |
| 39 | 86.8 | 6.7 | 2985  | 1 X32266 | Human nuclear rece |
| 40 | 86.8 | 6.7 | 2987  | 1 X32267 | Human nuclear rece |
| 41 | 86   | 6.6 | 291   | 1 022192 | Tetranucleotide re |
| 42 | 86   | 6.6 | 291   | 1 022787 | Microsatellite rep |
| 43 | 86   | 6.6 | 291   | 1 057810 | Repeat polymorphis |
| 44 | 86   | 6.6 | 291   | 1 V15526 | Polymorphic region |
| 45 | 85.8 | 6.6 | 2223  | 1 Q80908 | Plasmodium falcipa |

## ALIGNMENTS

| RESULT | ID     | Query Match  | Score  | DB 1             | Length | DB 2 | Indels | Gaps |
|--------|--------|--|--------|------------------|--------|------|--------|------|
| 1      | X40056 | 99.2%  | 1288   | 1                | 1298   |      | 0      | 0    |
| 2      | X40056 | Best Local Similarity  | 100.0% | Pred. No. 2e-224 |        |      |        |      |
| 3      | X40056 | Matches 1298; Conservative   | 0      | Mismatches 0     |        |      |        |      |
| 4      | X40056 | Standard; DNA; 1298 BP.  |        |                  |        |      |        |      |
| 5      | X40056 | 02-JUL-1999 (first entry)  |        |                  |        |      |        |      |
| 6      | X40056 | Colon cancer associated gene.  |        |                  |        |      |        |      |
| 7      | X40056 | Cancer associated antigen; diagnosis; research; treatment; human;            |        |                  |        |      |        |      |
| 8      | X40056 | breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;      |        |                  |        |      |        |      |
| 9      | X40056 | prostate cancer; ss.   |        |                  |        |      |        |      |
| 10     | X40056 | Home sapiens.  |        |                  |        |      |        |      |
| 11     | X40056 | WO904265-A2.   |        |                  |        |      |        |      |
| 12     | X40056 | 28-JAN-1999.   |        |                  |        |      |        |      |
| 13     | X40056 | 15-JUL-1998; U14679.   |        |                  |        |      |        |      |
| 14     | X40056 | 22-JUN-1998; US-102322.  |        |                  |        |      |        |      |
| 15     | X40056 | 17-JUL-1997; US-896164.  |        |                  |        |      |        |      |
| 16     | X40056 | 10-OCT-1997; US-061599.  |        |                  |        |      |        |      |
| 17     | X40056 | 10-OCT-1997; US-061765.  |        |                  |        |      |        |      |
| 18     | X40056 | 10-OCT-1997; US-948705.  |        |                  |        |      |        |      |
| 19     | X40056 | 11-OCT-1997; GB-021697.  |        |                  |        |      |        |      |
| 20     | X40056 | (LUDWIG) LUDWIG INST CANCER RES.   |        |                  |        |      |        |      |
| 21     | X40056 | Chen Y, Gout I, Gure A, O'hare M, Obara Y, Old LJ,                           |        |                  |        |      |        |      |
| 22     | X40056 | Pfreundschuh M, Sahn U, Scanlan MJ, Stockert E,                              |        |                  |        |      |        |      |
| 23     | X40056 | Tureci O;  |        |                  |        |      |        |      |
| 24     | X40056 | WPI; 99-132448/11.   |        |                  |        |      |        |      |
| 25     | X40056 | New isolated cancer associated nucleic acids and polypeptides -              |        |                  |        |      |        |      |
| 26     | X40056 | PT Isolated using sera from cancer patients, used to develop products        |        |                  |        |      |        |      |
| 27     | X40056 | for the diagnosis, monitoring or treatment of cancers                        |        |                  |        |      |        |      |
| 28     | X40056 | PS Claim 67; Page 654-655; 787pp; English.                                   |        |                  |        |      |        |      |
| 29     | X40056 | CC The invention relates to a method for diagnosing a disorder characterised |        |                  |        |      |        |      |
| 30     | X40056 | CC by expression of a human cancer associated antigen precursor coded for by |        |                  |        |      |        |      |
| 31     | X40056 | CC a nucleic acid molecule (NMW). The method comprises: (a) contacting a     |        |                  |        |      |        |      |
| 32     | X40056 | CC biological sample isolated from a subject with an agent that specifically |        |                  |        |      |        |      |
| 33     | X40056 | CC binds to the NMW, an expression product or a fragment of an expression    |        |                  |        |      |        |      |
| 34     | X40056 | CC product complexed with an HLA molecule; and (b) determining the           |        |                  |        |      |        |      |
| 35     | X40056 | CC interaction between the agent and the NMW or the expression product as a  |        |                  |        |      |        |      |
| 36     | X40056 | CC determination of the disorder. The products and methods can be used in    |        |                  |        |      |        |      |
| 37     | X40056 | CC the diagnosis, monitoring, research, or treatment of conditions           |        |                  |        |      |        |      |
| 38     | X40056 | CC characterised by the expression of various cancer associated antigens.    |        |                  |        |      |        |      |
| 39     | X40056 | CC The invention provides nucleic acid sequences and encoded polypeptides    |        |                  |        |      |        |      |
| 40     | X40056 | CC which are cancer associated antigen precursors expressed in human breast  |        |                  |        |      |        |      |
| 41     | X40056 | CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and   |        |                  |        |      |        |      |
| 42     | X40056 | CC lung cancer.  |        |                  |        |      |        |      |
| 43     | X40056 | Sequence 1298 BP; 546 A; 189 C; 379 G; 174 T;                                |        |                  |        |      |        |      |

[illegible][illegible]

| Query Match  | Similarity       | Score   | DB 1   | Length |
|--------------|------------------|---|--------|--------|
| Best Local   | Similarity 57.3% | Pred. 1.3e-26;  |        |        |
| Matches 340; | Conservative     | 0; Mismatches 253;  | Indels | Gaps   |
|              |                  |   | 0      | 0      |
| QY           | 706              | tcgagctcctactatgagncgaagcgtcgcagagagagagaagaatcnaagtltaagatlc | 765    |        |
|              |                  |   |        |        |
| Db           | 3730             | TCGAGTTCGCAA          | 3789   |        |
| QY           | 766              | acaagctgtgaagaagaaagccaaagaagccctaaagaagtttgacgtcgcga         | 825    |        |
|              |                  |   |        |        |
| Db           | 3790             | AA          | 3849   |        |
| QY           | 826              | aggtatccagctgcgcgcactagaagaacgaagaagagaaagagagagaga           | 885    |        |
|              |                  |   |        |        |
| Db           | 3850             | AA          | 3909   |        |
| QY           | 886              | agaagaagacaaagagaaagaagaagaagagagagagaaagaaagaagagag          | 945    |        |
|              |                  |   |        |        |
| Db           | 3910             | AA          | 3969   |        |



[illegible]

| Query Match               | Best Local Similarity | Score              | DB 1 | Length          | 7372             |
|---------------------------|-----------------------|--------------------|------|-----------------|------------------|
| Matches 340: Conservative | 57.3%                 | Pred. NO. 1.3e-26; | 0;   | Mismatches 253; | Indels 0; Gaps 0 |

[illegible]

RESULT 4  
 X33180  
 ID X33180 standard; DNA; 7/97 BP.  
 AC X33180;  
 DT 25-JUN-1999 (first entry)  
 DE COMPOX virus bsr full length gene sequence.  
 KW COMPOX virus; bsr; viral vector; expression; apoptosis; resistance;  
 KW crm; bcl-2; bcl-xL; FLIP; SURVIVIN; IAP; ILP; adenovirus; cancer;  
 KW autoimmune disease; graft rejection reaction; inflammation;  
 KW inflammatory disease; ss.  
 OS COMPOX virus.  
 PN W09913073-A2.  
 PD 18-MAR-1999.  
 PF 07-SEP-1998; J04010.  
 PR 08-SEP-1997; JP-259235.  
 PI (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
 PI Hamada H;  
 DR WPI; 99-243728/20.  
 PT New apoptosis-resistant virus-sensitive cell  
 PS Example 1; Page 34-38; 51pp; English.  
 CC The present invention describes an apoptosis-resistant virus-sensitive  
 CC cell line into which an apoptosis resistance gene has been introduced.  
 CC The recombinant viruses generated are capable of expressing apoptosis-  
 CC associated genes. These can then be used in a variety of diseases for  
 CC which the induction of apoptosis by gene transfer, or where the  
 CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
 CC are useful as vectors for gene therapy which can be applied to cancer  
 CC therapy for destroying cancer cells selectively, the treatment of  
 CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
 CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
 CC encountered the problem where if an adenovirus vector capable of

CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the compox virus bar gene which  
CC is used in an example from the present invention.  
SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T;

Query Match 14.7%; Score 191.2; DB 1; Length 7797;  
Best Local Similarity 57.3%; Pred. No. 1.3e-26;  
Matches 340; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 706 tgcagctctactatgagccagagctcgagagagagaaatcnaagttcaaatc 765  
DB 4883 TGCAGTTCGCAA 4942  
QY 766 acaagtcgtgaagaaagagcccaagagccctaaagagttgagcagctgcgga 825  
DB 4943 AA 5002  
QY 826 aggttaatccagctgcgcactagaagacgaagaaagagagagagagagaa 885  
DB 5003 AA 5062  
QY 886 agaagaagaacagagagagagaaagagagagagagagagagagagagagag 945  
DB 5063 AA 5122  
QY 946 aggaagaag 1005  
DB 5123 AA 5182  
QY 1006 gaactaagaag 1065  
DB 5183 AA 5242  
QY 1066 aagaagaag 1125  
DB 5243 AA 5302  
QY 1126 gaaagaagaatcaagacnagaagagagagagagagagagagagagagag 1185  
DB 5303 AA 5362  
QY 1186 agaagaagaag 1245  
DB 5363 AA 5422  
QY 1246 agcggaagaagaagaagtataagaagagagagagagagagagagagaa 1298  
DB 5423 AA 5475

RESULT 5  
X33184  
ID X33184 standard; DNA; 7996 BP.  
AC X33184;  
DT 25-JUN-1999 (first entry)  
DE Base sequence of the plasmid pRX-Bcl 2-1-hCD 25.  
KW Compo virus; bar; viral vector; expression; apoptosis; resistance;  
KW ctmA; bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;  
KW autoimmune disease; graft rejection reaction; inflammation;  
KW inflammatory disease; ss.  
OS Synthetic.  
OS Homo sapiens.  
PN W09913073-A2.  
PD 18-MAR-1999.  
PF 07-SEP-1998; J04010.  
PR 08-SEP-1997; JP-259235.

PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.  
PI Hamada H;  
DR WPI: 99-243728/20.  
PT New apoptosis-resistant virus-sensitive cell  
PS Example 3; Page 46-49; 51pp; English.  
CC The present invention describes an apoptosis-resistant virus-sensitive  
CC cell line into which an apoptosis resistance gene has been introduced.  
CC The recombinant viruses generated are capable of expressing apoptosis-  
CC associated genes. These can then be used in a variety of diseases for  
CC which the induction of apoptosis by gene transfer, or where the  
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses  
CC are useful as vectors for gene therapy which can be applied to cancer  
CC therapy for destroying cancer cells selectively, the treatment of  
CC autoimmune diseases and graft rejection reaction, and apoptosis induction  
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have  
CC encountered the problem where if an adenovirus vector capable of  
CC expressing an apoptosis-associated gene is introduced into animal cells,  
CC the cells producing the virus will be destroyed because the period of  
CC time required to induce cell death by apoptosis is shorter than that  
CC required to replicate and produce the virus, resulting in failure to  
CC obtain a recombinant virus having the integrated apoptosis-associated  
CC gene. In this invention an apoptosis-resistant 293 cell line (having an  
CC apoptosis resistant gene introduced) is established and overcomes the  
CC problem. The present sequence represents the base sequence of the  
CC plasmid pRX-Bcl 2-1-hCD 25, which contains the human Bcl-2 gene, and  
CC is used in an example from the present invention.  
SQ Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T;

Query Match 14.7%; Score 191.2; DB 1; Length 7996;  
Best Local Similarity 57.3%; Pred. No. 1.3e-26;  
Matches 340; Conservative 0; Mismatches 253; Indels 0; Gaps 0;

QY 706 tgcagctctactatgagccagagctcgagagagagaaatcnaagttcaaatc 765  
DB 5082 TGCAGTTCGCAA 5141  
QY 766 acaagtcgtgaagaaagagcccaagagagcttaagagagttgagcagctgcgga 825  
DB 5142 AA 5201  
QY 826 aggttaatccagctgcgcactagaagacgaagaaagagagagagagagagaa 885  
DB 5202 AA 5261  
QY 886 agaagaagaacag 945  
DB 5262 AA 5321  
QY 946 agaagaagaag 1005  
DB 5322 AA 5381  
QY 1006 gaactaagaag 1065  
DB 5382 AA 5441  
QY 1066 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1125  
DB 5442 AA 5501  
QY 1126 gaaagaagaatcaagacnagaagagagagagagagagagagagagagagag 1185  
DB 5502 AA 5561  
QY 1186 agaagaagaag 1245  
DB 5562 AA 5621  
QY 1246 agcggaagaagaagaagtataagaagagagagagagagagagagagagaa 1298  
DB 5622 AA 5674

RESULT 6  
ID 087587 standard; DNA: 1686 BP.  
AC 087587;  
DE 19-DEC-1995 (first entry)  
KW DNA encoding Leucocytozoan protozoa structural protein epitope.  
OS Leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;  
KM Leucocytozoan protozoa; treatment; ss.  
LN Leucocytozoan protozoa sp.  
PN J07089995-A.  
PD 04-APR-1995.  
PF 10-SEP-1993; 226078.  
PR 10-SEP-1993; JP-226078.  
PA (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.  
PM (NISS-) NISSEIKEN KK.  
DR WPI; 95-167252/22.  
P-PSDB; R70491.  
PT Immune inducing polypeptide against leucocytozoan protozoa - useful  
in production of vaccines for treatment of leucocytozoanosis in  
fowl.  
PS Claim 1; Page 12-14; 20pp; Japanese.  
CC 087587-89 encode polypeptides having a whole or partial epitope of a  
structural protein of Leucocytozoan protozoa (see R70491.93). The  
CC polypeptides and DNA encoding them are useful in the production of  
CC vaccines for the treatment of leucocytozoanosis of fowl.  
SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T;

Query Match 11.4%; Score 148.2; DB 1; Length 1686;  
Best Local Similarity 59.6%; Pred. No. 6.7e-19;  
Matches 317; Conservative 0; Mismatches 211; Indels 4; Gaps 4;

QY 765 aagtcgtgaaagaagaagccaaagaagccctaaagaagtttgacgctgcggaag 828  
DB 101 AAGGAAGAAATGACGAAATGAAAAAGAAAGAACAAAGAAAGAAAGAAAGAAAG 160  
QY 829 ttaatccgcgtccgacacagaaagcgaagaagaagaaagaaagaaagaaag 887  
DB 161 AACCAAGAACAAAGAAATCGTAGAAGAACCAAGAACCAAGAACCAAGAACCAAG 220  
QY 888 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 947  
DB 221 ATGGAAGAACAAAGAAAGAA-AAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 279  
QY 948 gaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1006  
DB 280 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 339  
QY 1007 aactaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1066  
DB 340 GATGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 399  
QY 1067 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1126  
DB 400 CAAATGAGATGAAACAAATGAAAGTGAACAAATGAAAGAAAGAAAGAAAGAAAG 459  
QY 1127 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1186  
DB 460 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 519  
QY 1187 gaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1245  
DB 520 GAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 579  
QY 1246 aacggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1297  
DB 580 AATGAAGATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 631

RESULT 7  
T91902  
ID T91902 standard; DNA: 4000 BP.

AC T91902;  
DE 07-MAY-1998 (first entry)  
KW Mannose-1-phosphate transferase gene MNNA.  
KW Mannose-1-phosphate transferase; MNNA gene; enzyme; yeast; regulatory gene;  
human; high mannose type neutral saccharide chain; ss.  
OS Saccharomyces cerevisiae.  
FH Key location/Qualifiers  
FT CDS 418..3953  
FT /tag= a  
FT /product= MNNA

PN J09266792-A.  
PD 14-OCT-1997.  
PF 29-MAR-1996; 075667.  
PR 29-MAR-1996; JP-075667.  
PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.  
PM WPI; 97-553460/51.  
DR P-PSDB; W30763.  
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast -  
useful for high mannose type neutral saccharide chain production  
PS Claim 2; Page 14-17; 23pp; Japanese.  
CC This sequence represents the gene of the invention, designated MNNA,  
CC encoding a protein which positively regulates mannose-1-phosphate  
CC transfer in yeast. The gene is useful for the preparation of human high  
CC mannose type neutral saccharide chain.  
SQ Sequence 4000 BP; 1453 A; 750 C; 770 G; 1027 T;

Query Match 11.2%; Score 145; DB 1; Length 4000;  
Best Local Similarity 65.9%; Pred. No. 2.7e-18;  
Matches 274; Conservative 0; Mismatches 128; Indels 14; Gaps 4;

QY 882 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 941  
DB 3527 AGAAACCGAGCGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3586  
QY 942 gaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 999  
DB 3587 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3646  
QY 1000 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1059  
DB 3647 AAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3705  
QY 1060 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1119  
DB 3706 GGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3759  
QY 1120 aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1179  
DB 3760 GATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3819  
QY 1180 agaaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 1239  
DB 3820 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3879  
QY 1240 gtagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1295  
DB 3880 GAAGCGAGAGG-----AGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3930

RESULT 8  
T05868  
ID T05868 standard; DNA: 3399 BP.  
AC T05868;  
DE 14-AUG-1996 (first entry)  
KW Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.  
KW Chicken leucocytozoan; immunogen; recombinant vaccine; protection;  
immunisation; vaccination; ss.  
OS Chicken leucocytozoan.  
FH Key location/Qualifiers  
FT cds 1..3399  
FT /tag= a  
FT misc\_feature 1150..3218

|                           |  |   |
|---------------------------|--|---|
| FT                        |  | /tag= b   |
| FT                        |  | /note= "fragment referred to in the claims, for |
| FT                        |  | use as insert in a recombinant vaccine          |
| FT                        |  | against chicken leucocytozoan disease"          |
| PN                        | J07284392-A.   |   |
| PD                        | 31-OCT-1995.   |   |
| PF                        | 19-APR-1994; JP-080643.  |   |
| PR                        | 19-APR-1994; JP-080643.  |   |
| PA                        | (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.                         |   |
| PP                        | (KITA) KITASATO KENKYUSHO SH.  |   |
| DR                        | WPI; 96/006311/01.   |   |
| DR                        | P-PEDB; R97866.  |   |
| PT                        | Chicken leucocytozoan immunogenic protein - used in a recombinant        |   |
| PS                        | vaccine against chicken leucocytozoan disease                            |   |
| CC                        | Claim 6; Page 6-9; 35pp; Japanese.                                       |   |
| CC                        | T05868 encodes a chicken leucocytozoan immunogenetic protein, this DNA   |   |
| CC                        | or a fragment of it can be used in a recombinant vaccine to immunise     |   |
| CC                        | against chicken leucocytozoan disease. The DNA is used in a vector       |   |
| CC                        | and operatively linked to an expression regulatory sequence as in        |   |
| CC                        | standard practice.   |   |
| SQ                        | Sequence 3399 B; 1577 A; 508 C; 798 G; 516 T;                            |   |
| <hr/>                     |  |   |
| Query Match               | 11.0%; Score 142.2; DB 1:  | Length 3399g;                                   |
| Best Local Similarity     | 57.6%;   | Pred. No. 8.5e-18;                              |
| Matches 328; Conservative | 0; Mismatches 232;   | Indels 9; Gaps                                  |
| OY                        | 739 aagaagaaatcnaaagttaaagtaacaaaagtcgtgaagaagaagaaggccaaagaag           | 798   |
| DG                        |  |   |
| DG                        | 2595 AGAAGAGAAGAAAGAAAGAAAAGTACATCAACAAGAGAAAGAAAAGTAACCATGAAGA          | 2654  |
| OY                        | 799 ccctaaagaagtttgcagcgtcggaagtgattccagctgcgcactagaagaagcaa             | 858   |
| DG                        |  |   |
| DG                        | 2655 AGAAGAAAAAGGTAAACATCGAAGAAAGAAAGAAAAGTAAATCATGAAGAAAGAAAGAAAGA      | 2714  |
| OY                        | 859 gaaaagggaaaaaggagggaggaagaagaagaacaaggagaagaagaagaagaag              | 918   |
| DG                        |  |   |
| DG                        | 2715 AGAACATGGGGAAGAGAGAAAGAAAGAAAGAAAGAAAGAGAAAGAAAGAAAGAAAGA           | 2774  |
| OY                        | 919 gagagaaggaaga-----aagaagggaagggaagaagaaggagaagaagaag                 | 973   |
| DG                        |  |   |
| DG                        | 2775 TGAGCAGACGAGAGAGAGAAAGAAAGAAAGAAAGTATGAGAGAGAGAGAGAAAGAAAGAAATGA    | 2834  |
| OY                        | 974 agaaggaagaagaagaagaagaagaaga--agaagaactaagaagaagaagaagaata           | 1032  |
| DG                        |  |   |
| DG                        | 2835 GGAAAGACAGACAGAGAAAGAAATAAGAGAGAAAGAAAGAGAAAGAAAGACATGAAGA          | 2894  |
| OY                        | 1033 agaaaggaaagaagaagaagaagaagtcnaaagaagaagaagaagaagaagaagaag           | 1092  |
| DG                        |  |   |
| DG                        | 2895 AGAAATTAAACACTTGAGAGAGAGAAAGAAAGAAAGTTAACACTTGAGAGAGAGAAAGAAAGTAAAC | 2954  |
| OY                        | 1093 gaagaacctnagaagaagaagaaggagggaagaagaagaagaagaagaagaagaaga           | 1152  |
| DG                        |  |   |
| DG                        | 2955 ACATGAAAGAGAGAGAAAGTATGTCATATGAGAGAGAGAGAGAGAGAAAGTAACTATGAGAGA     | 3014  |
| OY                        | 1153 ggaagaagaagaataagaagaagaagaagaagaag-aagaagaagaagaagaagaaga          | 1211  |
| DG                        |  |   |
| DG                        | 3015 AGAAGAAAAAGTAAACATGAGAGAGAGAGAAAGAAAGTAAACATGAGAGAGAGAGAGAGAGAGT    | 3074  |
| OY                        | 1212 gaaaaaagaaga--aaaaaggaagaagaagttagaagaagcggaagaagaagaagaagtcat      | 1269  |
| DG                        |  |   |
| DG                        | 3075 AACCATATGAGAGAGAGAGAGAAAGAAAGTAAACATGAGAGAGAGAGAGAGAGAAAGTAAATGAA   | 3134  |
| OY                        | 1270 aagaaggaagaagaagaagaagaagaaga                                       | 1298  |
| DG                        |  |   |
| DG                        | 3135 CAAGAGAGAGATATGAAGTATATAAAAAAAAA                                    | 3163  |
| <hr/>                     |  |   |
| RESULT                    | 9  | X02998  |
| ID                        | X02998 standard; DNA; 11901 BP.  |   |
| AC                        | X02998;  |   |
| DT                        | 22-JUN-1999 (first entry)  |   |

|    |   |
|----|---|
| DE | Human IL-1ra BAC contiguous DNA sequence 43.                              |
| KW | Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;  |
| KM | Interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;  |
| KW | chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;     |
| KM | growth factors; treatment; IL-1 receptor complex; BAC; ss.                |
| OS | Homo sapiens.   |
| PN | M0906426-R1.  |
| PD | 11-FEB-1999.  |
| PF | 03-JUL-1998; 016102.  |
| PR | 02-AUG-1998; US-091650.   |
| PR | 04-AUG-1997; US-054646.   |
| PA | (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.                                   |
| PI | Pan Y;  |
| PT | WPI; 99-153692/13.  |
| PT | New isolated nucleic acid encoding the new human cytokine Tango-77 -      |
| PS | used to inhibit inflammation and to screen for specific modulators        |
| PP | Example 5; Figure 3; 226bp; English.                                      |
| CC | X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences     |
| CC | containing alternatively spliced forms of human IL-1ra. Such fragments    |
| CC | are used in the method of the invention which describes the isolation of  |
| CC | a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a   |
| CC | member of the cytokine superfamily that is expected to inhibit            |
| CC | inflammation by binding to the interleukin-1 receptor (IL-1R). It may     |
| CC | also bind to a new receptor so could regulate other cellular processes    |
| CC | associated with acute or chronic inflammation, e.g. asthma, chronic       |
| CC | myelogenous leukemia, rheumatoid arthritis, psoriasis and inflammatory    |
| CC | bowel disease. It may also induce or suppress interleukins, cytokines and |
| CC | growth factors. Modulators of this protein are used to treat or prevent   |
| CC | conditions associated with abnormal levels of inflammation, or activity   |
| CC | of IL-1 or its receptor complex.  |
| SQ | Sequence 11901 BP; 3552 A; 2724 C; 2438 G; 3179 T;                        |

| Query Match           | 9.7%   | Score 125.6;      | DB 1;           | Length 11901; |
|-----------------------|--|-------------------|-----------------|---------------|
| Best Local Similarity | 61.1%;   | Pred No. 9.2e-15; |                 |               |
| Matches 218;          | Conservative   | 0;                | Mismatches 137; | Indels 2;     |
|                       |  |                   | Gaps            | 1             |
| OY 853                | aacgaagaaagaagaaagaggagagaaagaaagaaacaaggaagaaag         | 912               |                 |               |
|                       |  |                   |                 |               |
| Db 6414               | ACCAGAAAACAAGCGACAGAGAGAGAGAGAAGGAAGGAAAGAAAGAAAGAAAG    | 6473              |                 |               |
| OY 913                | aagaaaggagaaagagaaagaaagaagaagaaagaaagaaagaaagaaag       | 972               |                 |               |
|                       |  |                   |                 |               |
| Db 6474               | AAAGAAAGAAATGCAGAGGAAGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAG   | 6533              |                 |               |
| OY 973                | gagaagaaagaaagaaagaaagaaagaagaactaagaagaagagagaaagata    | 1032              |                 |               |
|                       |  |                   |                 |               |
| Db 6534               | AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG     | 6593              |                 |               |
| OY 1033               | agaaggaagaaagaaagaaagaaagttaagaagaagaagaagaagaaagaaag--  | 1091              |                 |               |
|                       |  |                   |                 |               |
| Db 6594               | ACGAGGAAGGAAGGAAGAAAGAAAGCAAGCAAGCAGGAAGGAAGGAAGGAAGGA   | 6653              |                 |               |
| OY 1091               | aggaagaactnagaagaagaagagaggaagaagaagaagaatlaaggaacnagaag | 1150              |                 |               |
|                       |  |                   |                 |               |
| Db 6654               | ACGAGGAAGGAAGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG    | 6713              |                 |               |
| OY 1151               | aaggaagaagaagaatagaagaagaagaagaagaagaagaagaagaagaagaaga  | 1207              |                 |               |
|                       |  |                   |                 |               |
| Db 6714               | AAGAGCAAAGAAAGAAAGGAAGGAGAGGAAGGAAGGAAGAAAGAAAGCAAGAAAGA | 6770              |                 |               |

|   |
|---|
| RESULT 10   |
| V62142/c  |
| ID V62142 standard; DNA: 2056 BP.                               |
| AC V62142:  |
| DT 07-DEC-1998 (first entry)                                    |
| DR HSV-2 strain SB5 Contlig ID 82 DNA sequence.                 |
| KM HSV-2 strain SB5: immunological response induction; therapy; |
| KW antiviral identification; viral protein inhibitor; ss.       |
| OS Herpes simplex virus type 2.                                 |
| FH key Location/Qualifiers                                      |
| CDS 761..1150   |







Dh 8270 AGTAAGTGAATAAAGGATGCTTAAGCAAAATAATATTAAAGAAAAGAGAG 8211  
Qy 918 gggagagaggaagaaagagagagagagagagagagagagagagagagagag 977  
Dh 8210 AGAAGTACTAAGAGAGCGGAGAGACTAGACAGACAGATGATGAAGAA--GATAGATAGA 8153  
Qy 978 ggaagaaagaaag 1037  
Dh 8152 GTAACTGACGTAGAGAAAGAAATGAAATACGACACATATTAAAGAAAGGAAAGATAGCT 8093  
Qy 1038 gaaagaaagaaagaaag 1097  
Dh 8092 GACAGTACTAGTACGAGAGAGAGAAAGTAAAGAGAGAGAGAGAGAGATATAGAAAGAAAG 8033  
Qy 1098 actag 1155  
Dh 8032 AGAAAGGAAAGAGAGAAAG 7973  
Qy 1156 gaagaaag 1215  
Dh 7972 TAG 7913  
Qy 1216 aaggaag 1275  
Dh 7912 GAGAAAG 7853  
Qy 1276 gaagaaag 1298  
Dh 7852 GAGGAGAAAG 7830

## RESULT 14

X23519  
ID X23519 standard; DNA: 44453 BP.  
AC X23519;  
DT 23-JUN-1999 (first entry)  
DE Human kidney aminopeptidase P genomic DNA fragment 3.  
KW Aminopeptidase; human; Amp; gene therapy; treatment; Amp-deficiency;  
KW prenatal diagnosis; angioedema; antihypertensive agent; atherosclerosis;  
KW arterial stenosis; industrial protein feed; malabsorption syndrome;  
KW proteolaceous waste degradation; additive; immunohistochemistry; ss.  
OS Homo sapiens.  
PN W09911799-A2.  
PD 11-MAR-1999.  
PF 02-SEP-1998; U18426.  
PR 02-SEP-1997; US-057854.  
PA (MEDI-) MEDICAL COLLEGE GEORGIA RES INST.  
PI Ryan JM, Sprinkle TJC, Venema RC;  
DR WPI: 99-205193/17.  
PT Nucleic acid encoding human aminopeptidase P  
PS Claim 13; Page 139-165; 201pp; English.  
CC This invention describes the isolation of a novel human aminopeptidase P  
CC (AMP). This protein is used to produce recombinant Amp and can be used  
CC for gene therapy for treating Amp-deficiency conditions. Its fragments  
CC are used as primers and probes to identify patients with homozygous and  
CC heterozygous Amp deficiency, including prenatal diagnosis (patients  
CC heterozygous in Amp are at risk of developing angioedema if treated with  
CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors  
CC in cases of excessive Amp expression. The product of the invention is  
CC also used to identify Amp-expressing sequences in other animals and to  
CC generate transgenic animals, and comparisons of genomic sequences are  
CC used to detect mutations. Amp inhibitors are potentially useful as  
CC antihypertensive agents and to prevent or treat arterial (re)stenosis  
CC or atherosclerosis. The structure of Amp is used to design synthetic  
CC substrates, e.g. for use in Amp assays. Amp, which hydrolyzes N-terminal  
CC imido bonds, can be used to degrade industrial protein feeds to free  
CC amino acids, to degrade proteinaceous wastes, as additives in enzyme  
CC formulations used to treat malabsorption syndrome and for studying its  
CC biological role. Antibodies against Amp are used in immunohistochemical  
CC methods to study Amp distribution.  
SQ Sequence 44453 BP; 13034 A; 10110 C; 10292 G; 11017 T;

Query Match 9.0%; Score 116.2; DB 1; Length 44453;  
Best Local Similarity 59.7%; Pred. No. 5.1e-13;  
Matches 210; Conservative 0; Mismatches 141; Indels 1; Gaps 1;  
Qy 865 aggaag 924  
Dh 20817 AGAAAGAAAGAAAG 20876  
Qy 925 ggaag 984  
Dh 20877 GGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 20936  
Qy 985 gaaag 1044  
Dh 20937 GAGAGAAAG 20996  
Qy 1045 gaaagaaag 1104  
Dh 20997 AAGACAGAGAA--AGAAAGGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 21055  
Qy 1105 agaag 1164  
Dh 21056 AAGAAAGAGAGAAAG 21115  
Qy 1165 ataag 1216  
Dh 21116 GAAAG 21167

## RESULT 15

V15422  
ID V15422 standard; DNA: 29392 BP.  
AC V15422;  
DT 11-JUN-1998 (first entry)  
DE Mouse poly Ig receptor protein gene.  
KW Mouse; poly Ig receptor protein; pIgr protein; pig; deficiency;  
KW knockout mouse; disease model; ds.  
OS Mus sp.  
PN J10057066-A.  
PD 03-MAR-1998.  
PF 19-AUG-1996; 217154.  
PR 19-AUG-1996; JP-217154.  
PA (HONS) YAKULT HONSHA KK.  
DR WPI: 98-254323/23.  
PT Mouse pig receptor protein gene - used for preparing gene knockout  
PT mice, useful for study of human poly Ig receptor protein deficiency  
PS Claim 1; Page 4-14; 18pp; Japanese.  
CC The present sequence represents the mouse poly Ig receptor protein  
CC gene, which has a 29392 bp sequence. The new gene can be used to  
CC produce a gene knockout mouse, useful as a disease model of human  
CC poly Ig receptor protein deficiency.  
SQ Sequence 29392 BP; 8318 A; 6747 C; 6514 G; 7813 T;

Query Match 8.8%; Score 114.6; DB 1; Length 29392;  
Best Local Similarity 61.0%; Pred. No. 9.5e-13;  
Matches 183; Conservative 0; Mismatches 117; Indels 0; Gaps 0;  
Qy 916 aaggaag 975  
Dh 28456 AAAG 28515  
Qy 976 aaggaag 1035  
Dh 28516 AAAAG 28575  
Qy 1036 aggaag 1095  
Dh 28576 AGAAAG 28635  
Qy 1096 gaacnaag 1155  
Dh 28636 AGAAG 28695







|    |                       |  |                    |           |              |   |
|----|-----------------------|--|--------------------|-----------|--------------|---|
|    | Query Match           | 3.6%   | Score 81.4         | DB 1,     | length 7218: |   |
|    | Best Local Similarity | 4.7%   | Pred. No. 1.le-11: |           |              |   |
|    | Matches 19;           | Conservative 246;  | Mismatches 142;    | Indels 0; | Gaps 0;      |   |
| QY | 1032                  | gcagcgcgacgaagcttcattcatgcagaagcgcgttgcgatgataatcccaaatccctcca   | 1091               |           |              |   |
|    | :                     | :  | :                  | :         | :            | : |
| Db | 1438                  | GTACRRR    | 1379               |           |              |   |
| QY | 1092                  | ggagcagcagcgagatcggcgcgcaaaagagaagaataatgccccaaggcgcgcagaga      | 1151               |           |              |   |
| Db | 1378                  | RR       | 1319               |           |              |   |
| QY | 1152                  | aaatgaagatatccggaagagatgtaaacagatctaagggaggaagaagttaaaga         | 1211               |           |              |   |
| Db | 1318                  | RR       | 1259               |           |              |   |
| QY | 1212                  | gcaatcgtggaagaagacttggtgcctcaagaagacagctactcttgtcctaaaacatcaatgc | 1271               |           |              |   |
| Db | 1258                  | RR       | 1199               |           |              |   |
| QY | 1272                  | tgaagttacaaccagaccttcgcgaagccaagatgatcagggagtggaacctgagct        | 1331               |           |              |   |
| Db | 1198                  | RR       | 1139               |           |              |   |
| QY | 1332                  | cgaagcccagatgactcgtgatcgtgagcagcaggaagacagggagacgaatcttcogga     | 1391               |           |              |   |
| Db | 1138                  | RR       | 1079               |           |              |   |
| QY | 1392                  | atacgaagaagccttgacacctactatgctatgtcacccccagacaga                 | 1438               |           |              |   |
| Db | 1078                  | RRRRRRRRRRATCGAAGCTCCCTGCACTGCACCTGACGCCAAAGTCTCGA               | 1032               |           |              |   |

RESULT 2  
US-08-728-323A-1  
Sequence 1, Application US/08728323A  
Patent No. 5948676  
GENERAL INFORMATION:  
APPLICANT: Chan, Yuen  
APPLICANT: Bohenzky, Roy A.  
APPLICANT: Russo, James J.  
APPLICANT: Edelman, Isidore S.  
APPLICANT: Moore, Patrick S.  
TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
TITLE OF INVENTION: Encoding Same And Uses Thereof  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooper & Dunham LLP  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/728,323A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 0575/52266/JFW/MSC/SKS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-378-0400  
TELEFAX: 212-391-0525  
INFORMATION FOR SEO ID NO: 1  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3489 base pairs

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
;
US-08-728-323A-1

```

| Query Match           | 3.1%   | Score 68.4      | DB 4           | length 3489 |
|-----------------------|--|-----------------|----------------|-------------|
| Best Local Similarity | 47.6%  | Pred. No. 2e-08 |                |             |
| Matches 201           | Conservative   | 0               | Mismatches 221 | Indels 0    |
| QY 979                | gagctgttcctcgtacagacacgagcgagcgctgtgcagaaagcgcgacgctgtgctcagcgcg     | 1038            |                |             |
| Db 2182               | GAGCAGCAGACGACGAGCATGTAGCGACGACAGAGATGAGCAGCAGCAGCAGCAGATGACGAG      | 2241            |                |             |
| QY 1039               | caggaagcttcctatgcagaagcgcgctgcggtatgagttccaaagaatcctccagaagcag       | 1098            |                |             |
| Db 2242               | CAGCAGCAGAGATGAACACGGAGCGACGAGGAGACACGAGACACGAGAGCAGCAGGAGCAG        | 2301            |                |             |
| QY 1099               | caggagatctggacgcgcacaaagagaaagaatctgccaaagagcagcagcagagaaatctg       | 1158            |                |             |
| Db 2302               | GAGTTAGGAGGACGAGGACGAGGAGTTAGAGGATCAGAGACAGGAGTTAGAGGACGAGGAG        | 2361            |                |             |
| QY 1159               | agatacccggaagagatgtgaacacgatctgtatagagagagaaagtaagtaagaacaatcgtg     | 1218            |                |             |
| Db 2362               | CAGGAGTTTAAAGGAGCGACGAGGACGAGGTTTAAAGGAGACGAGGACGAGGAGTTTAAAGGAGCAG  | 2421            |                |             |
| QY 1219               | gaagaagacctctggggcctcaaaagagacacgctactccttcctcaaaacatcatcctgtcaggtat | 1278            |                |             |
| Db 2422               | GAGCAGGAGTTTAAAGGAGGACGAGGAGGAGGAGTTTAAAGGACGAGGACGAGGAGTTTAAAGGAG   | 2481            |                |             |
| QY 1279               | caaccagctaccctctgcgaagccaaagtatgatcagggagctgtgaacctgtgagctcgagccc    | 1338            |                |             |
| Db 2482               | CAGGAGCAGAGAGTTTAAAGGAGGACGAGGACGAGGATTTAGAGAGCAGCAGAGGTTGCAACAGCA   | 2541            |                |             |
| QY 1339               | gcagatgacctgatatgtgaagcagcgagggagcaagggacagagatcttcgggaatatgtg       | 1398            |                |             |
| Db 2542               | GAGCAGGAGGTTGGAAGGCAAGGACGAGGACGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAG     | 2601            |                |             |
| QY 1399               | ga 1400  |                 |                |             |
| Db 2602               | GA 2603  |                 |                |             |

RESULT 3  
 US-08-770-379-20/C  
 : Sequence 20, Application US/08770379  
 : Patent No. 5849564  
 : GENERAL INFORMATION:  
 : APPLICANT: Chang, Yuan  
 : APPLICANT: Bohenzky, Roy A.  
 : APPLICANT: Russo, James J.  
 : APPLICANT: Edelman, Isidore S.  
 : APPLICANT: Moore, Patrick S.  
 : TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
 : TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
 : NUMBER OF SEQUENCES: 20  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Cooper & Dunham LLP  
 : STREET: 1185 Avenue of the Americas  
 : CITY: New York  
 : STATE: New York  
 : COUNTRY: U.S.A.  
 : ZIP: 10036  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: PatentIn Release 1.0, Version #1.30  
 : CURRENT APPLICATION DATA:



```

: FILE REFERENCE: 0867/ID0903US1
: CURRENT APPLICATION NUMBER: US/09/130,114
: CURRENT FILING DATE: 1998-08-06
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FASTSEQ for Windows Version 3.0.
: SEQ ID NO 1
: LENGTH: 5452
: TYPE: DNA
: ORGANISM: YEBNA
: US-09-130-114-1

```

|                           |        |                    |           |              |
|---------------------------|--------|--------------------|-----------|--------------|
| Query Match               | 2.6%;  | Score 58;          | DB 4;     | Length 5452; |
| Best Local Similarity     | 46.9%; | Pred. No. 1.2e-05; |           |              |
| Matches 181; Conservative | 0;     | Mismatches 205;    | Indels 0; | Gaps 0;      |

|    |      |   |                              |      |
|----|------|---|------------------------------|------|
| QY | 997  | cggagcgagcttgcgcgaagggcgccagcttga                                 | ctgcacgcgcgaagcttctcatgacg   | 1056 |
|    |      |   |                              |      |
| Db | 1835 | CAGGAGCAGGACGAGGGCGACGAGCAGAGGGCGCAGGAGGGCGAGGAGCGACG             |                              | 1776 |
| QY | 1057 | aagcgcgtctgcgtatgtagtccaaacaagatctct                              | ccaaggaacgaagaaatgtagcgcgcaa | 1116 |
|    |      |   |                              |      |
| Db | 1775 | GAGGGCGAGGAGCGAGGAGGAGGGCGACGAGGGCGCAGGAGGAGGAGGGCGAG             |                              | 1716 |
| QY | 1117 | agcgagaaagaatattgcgccagaaagcgacagagga                             | aaatgtagataccggaagagatg      | 1176 |
|    |      |   |                              |      |
| Db | 1715 | GAGCAGGAGGGCGCAGGAGCGAGGAGGGCGACAGACAGAGAGGGCGACGAGAGCGCAGGACGACG |                              | 1656 |
| QY | 1177 | gaacagatttagagagagagagaagatttaagaagcaat                           | tggaagaaagactcgtgcgtcca      | 1236 |
|    |      |   |                              |      |
| Db | 1655 | GAGGGCGCAGAGGGGGCGACGAGCGAGAGGGGCGAGAGGGGCGACGAGGAGGAGGGGCGAG     |                              | 1596 |
| QY | 1237 | aaggaaacagctactcttgccttaaacatcaactcgtct                           | gtgagtaacaccagtaacctcttcgc   | 1296 |
|    |      |   |                              |      |
| Db | 1595 | GAGGGCGAGGAGCGAGGAGGAGGGCGACGAGGCGCAGGAGGGCGCAGAGGGCGAG           |                              | 1536 |
| QY | 1297 | aagccaaagtattatcatcagggagttggaacctgagct                           | tgagcccgcaagatgacctgatatgga  | 1356 |
|    |      |   |                              |      |
| Db | 1535 | GAGCAGAGAGGGGCGAGGAGGGCGACGAGGAGGGGCGACGAGGAGGGCGACGAGCAGGAGGAG   |                              | 1476 |
| QY | 1357 | ggcacgagagagcagagagagagagagga                                     |                              | 1382 |
|    |      |   |                              |      |
| Db | 1475 | GGGCAGGAGCGAGGAGGGCGCAGGAGCGCA                                    |                              | 1450 |

RESULT 6  
US-07-884-811-15  
; Sequence 15, Application US/07884811

GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J., Lokker, Nathalie R.  
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/884, 811  
FILING DATE: 19920518  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:

```

? NAME: Dreger, Ginger R.
? REGISTRATION NUMBER: 33,055
? REFERENCE/DOCKET NUMBER: 755.1.1
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415/225-3316
? TELEFAX: 415/952-9881
?
? TEXT: 910/371-7168
? INFORMATION FOR SEQ ID NO: 15:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 10596 bases
? TYPE: NUCLEIC ACID
? STRANDEDNESS: single
? TOPOLOGY: linear
US-07-884-811-15
```

|                           |       |                    |           |               |
|---------------------------|-------|--------------------|-----------|---------------|
| Query Match               | 2.6%  | Score 58;          | DB 1;     | Length 10596; |
| Best Local Similarity     | 46.9% | Pred. No. 1.7e-05; |           |               |
| Matches 181; Conservative | 0;    | Mismatches 205;    | Indels 0; | Gaps 0;       |

[illegible]

```

1 RESULT 7
2 US-07-885-971-15
3 ; Sequence 15 Application US/07885971
4 Patent No. 5328837
5
6 GENERAL INFORMATION:
7
8 APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
9 TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS.
10 NUMBER OF SEQUENCES: 21
11 CORRESPONDENCE ADDRESS:
12
13 ADDRESSEE: Genentech, Inc.
14 STREET: 460 Point San Bruno Blvd
15 CITY: South San Francisco
16 STATE: California
17 COUNTRY: USA
18
19 ZIP: 94080
20
21 COMPUTER READABLE FORM:
22
23 MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
24
25 COMPUTER: IBM PC compatible
26 OPERATING SYSTEM: PC-DOS/MS-DOS
27 SOFTWARE: patin (Genentech)
28
29 CURRENT APPLICATION DATA:
30
31 APPLICATION NUMBER: US/07/885,971
32 FILING DATE: 19920518
33 CLASSIFICATION: 530

```

```

;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dieger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-885-971-15

Query Match      2.6%; Score 58; DB 1; Length 10596;
Best Local Similarity 46.9%; Pred. No. 1.7e-05;
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 997 cggagagcgtgcaagagcgcgccgagcgtgagctgagcggcagagagcttcctcatgcag 1056
    ||||| || || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2550 CAGGAGCAGAGAGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2609
QY 1057 aagcgcgtgctgagatgagtcacaaagatctccagagagcagagagatgagagcgcaa 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2610 GAGGGGAGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2669
QY 1117 aggaagaaaagaattgtcccaagagcagcagcagagagaaatgagagatcccgagagagatg 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2670 GACACAGAGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2729
QY 1177 gaacagattgagagagagagagaaagtttaagaagcaatgagagagagagagagctgagctca 1236
    || || || || || || || || || || || || || || || || || || || || || || || ||
DB 2730 GAGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2789
QY 1237 aaggaacagctactcttgcctaaacatcatcactgctgaggtacacccagtaacccttcgc 1296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2790 GAGGGGAGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2849
QY 1297 aagccaaagtatcatcagagagtggaacctgagctcgagcccgagatgagacttgatgga 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2850 GACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2909
QY 1357 ggcacgagagagcagagagagagcagga 1382
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2910 GGGCAGAGCAGAGAGGAGGAGGAGCAGAGCA 2935

RESULT 8
US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5547856
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCTYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)

```

```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/087,783A
; FILING DATE: 13-Jul-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/885971
; FILING DATE: 18-MAY-92
; ATTORNEY/AGENT INFORMATION:
; NAME: Marschang, Diane L.
; REGISTRATION NUMBER: 35,600
; REFERENCE/DOCKET NUMBER: P0755779P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-5416
; TELEFAX: 415/952-9881
; TELETYPE: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-087-783A-15

Query Match      2.6%; Score 58; DB 1; Length 10596;
Best Local Similarity 46.9%; Pred. No. 1.7e-05;
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 997 cggagagcgtgcaagagcgcgccgagcgtgagctgagcggcagagagcttcctcatgcag 1056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2550 CAGGAGCAGAGAGAGGAGGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2609
QY 1057 aagcgcgtgctgagatgagtcacaaagatctccagagagcagcagagatgagagcgcaa 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2610 GAGGGGAGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2669
QY 1117 aggaagaaaagaattgtcccaagagcagcagagagaaatgagagatcccgagagagatg 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2670 GACACAGAGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2729
QY 1177 gaacagattgagagagagagagaaagtttaagaagcaatgagagagagagagctgagctca 1236
    || || || || || || || || || || || || || || || || || || || || || || || ||
DB 2730 GAGGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2789
QY 1237 aaggaacagctactcttgcctaaacatcatcactgctgaggtacacccagtaacccttcgc 1296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2790 GAGGGGAGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2849
QY 1297 aagccaaagtatcatcagagagtggaacctgagctcgagcccgagatgagacttgatgga 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2850 GACGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2909
QY 1357 ggcacgagagagcagagagagagcagga 1382
    || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2910 GGGCAGAGCAGAGAGGAGGAGGAGCAGAGCA 2935

RESULT 9
US-08-194-088B-15
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCTYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California

```

```

; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-088B-15
```

```

Query Match          2.6%; Score 58; DB 1; Length 10596;
Best Local Similarity 46.9%; Pred. No. 1.7e-05;
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
```

```

QY 997 cggagcggctgagcagagcgcgagcgtgagctgacgcgagcagagcttcacagcag 1056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2550 CAGGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2609
QY 1057 aagcgctgctgagtgagtgccaaagatccctcagagcagagagatgagcgcgcaa 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2610 GAGGGGCGACGACGACGACGACGAGGCGCGACGACGACGACGACGACGAGGCGGCGAG 2669
QY 1117 aggaagaagaatctccagaagaagcagagagaatgagagatcaccggaagagagatg 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2670 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2729
QY 1177 gaacagattgtagaggaagaagaagtttaagaagaacatggaagaaagactgaggctca 1236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2730 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2789
QY 1237 aaggaacagctactcttgcttaaacatcactgctgaggtacaccagtaacccttcgc 1296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2790 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2849
QY 1297 aagccaagaatagatcaggaagtggaacctgagctcgagcccgacagatgacctgagtg 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2850 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2909
QY 1357 ggcacggaagcaggaaggaagcagga 1382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2910 GGGCAGGAGCAGGAGGAGGAGGAGGAGCA 2935

RESULT 10
US-08-194-087-15
; Sequence 15, Application US/08194087
; Patent No. 5879310
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCTYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
```

```

; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-194-087-15
```

```

Query Match          2.6%; Score 58; DB 3; Length 10596;
Best Local Similarity 46.9%; Pred. No. 1.7e-05;
Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0;
```

```

QY 997 cggagcggctgagcagagcgcgagcgtgagctgacgcgagcagagcttcacagcag 1056
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2550 CAGGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2609
QY 1057 aagcgctgctgagtgagtgccaaagatccctcagagcagagagatgagcgcgcaa 1116
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2610 GAGGGGCGACGACGACGACGACGAGGCGCGACGACGACGACGACGACGAGGCGGCGAG 2669
QY 1117 aggaagaagaatctccagaagaagcagagagaatgagagatcaccggaagagagatg 1176
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2670 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2729
QY 1177 gaacagattgtagaggaagaagaagtttaagaagaacatggaagaaagactgaggctca 1236
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2730 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2789
QY 1237 aaggaacagctactcttgcttaaacatcactgctgaggtacaccagtaacccttcgc 1296
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2790 GAGGGGCGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2849
QY 1297 aagccaagaatagatcaggaagtggaacctgagctcgagcccgacagatgacctgagtg 1356
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2850 GAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 2909
QY 1357 ggcacggaagcaggaaggaagcagga 1382
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2910 GGGCAGGAGCAGGAGGAGGAGGAGGAGCA 2935

RESULT 11
PCT-US93-04648-15
; Sequence 15, Application PC/TUS9304648
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie
```



```

14 RESULT
1 US-08-458-298-1
: Sequence 1, Application US/08458298
: Patent No. 5756677
:
: GENERAL INFORMATION:
:
: APPLICANT: Lewis, Randolph V.
: APPLICANT: Colgin, Mark
: TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
: TITLE OF INVENTION: Silk Proteins
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:

```

RESULT 15  
US-08-056-200-93  
; Sequence 93, Application US/08056200  
; Patent No. 5616500  
; GENERAL INFORMATION:  
; APPLICANT: Steinert, Peter M.  
; APPLICANT: Lee, Seung-Chul  
; APPLICANT: Kim, In-Gyu  
; APPLICANT: Chung, Soo-Il  
; APPLICANT: Park, Sang-Chul  
; TITLE OF INVENTION: Trichopyalin and Transglutaminase-3 and  
; TITLE OF INVENTION: Methods of Using Same  
; NUMBER OF SEQUENCES: 117  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Knobbe, Martens, Olson & Bear  
; STREET: 620 Newport Center Drive, Sixteenth Floor





**THIS PAGE BLANK (USPTO)**

|    |      |       |        |   |         |                     |
|----|------|-------|--------|---|---------|---------------------|
| 1  | 2236 | 100.0 | 2236   | 1 | X39644  | Renal cancer assoc  |
| 2  | 2236 | 100.0 | 2236   | 1 | X40057  | Colon cancer assoc  |
| 3  | 2173 | 97.2  | 2289   | 1 | X40061  | Colon cancer assoc  |
| 4  | 2078 | 92.9  | 2162   | 1 | X40058  | Colon cancer assoc  |
| 5  | 2053 | 91.8  | 2409   | 1 | X40062  | Colon cancer assoc  |
| 6  | 1306 | 58.4  | 1306   | 1 | X40060  | Colon cancer assoc  |
| c  | 68.4 | 3.1   | 32207  | 1 | V73805  | KSHV LTR DNA (nucl  |
| 7  | 68.4 | 3.1   | 137507 | 1 | V139947 | KSHV long unique c  |
| 8  | 58   | 2.6   | 799    | 1 | V55831  | Nucleotide sequenc  |
| 9  | 58   | 2.6   | 9600   | 1 | V21683  | Vector plasmid pcw  |
| 10 | 58   | 2.6   | 10596  | 1 | 051731  | Plasmid pCISEBON f  |
| 11 | 58   | 2.6   | 10596  | 1 | T40348  | Plasmid pCISEBON f  |
| 12 | 58   | 2.6   | 10596  | 1 | X15650  | Nucleotide sequenc  |
| 13 | 58   | 2.6   | 10596  | 1 | X15650  | FLGA insert stabili |
| 14 | 57   | 2.5   | 795    | 1 | V55830  | Antigen tc-7a gene  |
| 15 | 52   | 2.3   | 543    | 1 | 023092  | Mispl-containing F  |
| 16 | 49.4 | 2.2   | 2744   | 1 | 098470  | Mouse poly Ig rece  |
| 17 | 48.8 | 2.2   | 29392  | 1 | X54522  | Mannose-1-phosphat  |
| 18 | 47.6 | 2.1   | 4000   | 1 | T91902  | N. clavipes draglin |
| 19 | 46.8 | 2.1   | 2338   | 1 | 014183  | N. clavipes draglin |
| 20 | 46.8 | 2.1   | 2338   | 1 | V23249  | N. clavipes draglin |
| 21 | 46.4 | 2.1   | 2000   | 1 | N71065  | Gene encoding Plaa  |
| 22 | 46   | 2.1   | 243    | 1 | 037379  | Clone 2-9-used in   |
| 23 | 46   | 2.1   | 243    | 1 | 095004  | SSP 7.7.7.7.7.8     |
| 24 | 46   | 2.1   | 243    | 1 | V35842  | Synthetic storage   |
| 25 | 46   | 2.1   | 243    | 1 | V99539  | Synthetic lysine-r  |
| 26 | 46   | 2.1   | 1390   | 1 | V84524  | Human secreted p    |
| c  | 45.6 | 2.0   | 12001  | 1 | 076213  | HSV L/SF region, f  |
| 28 | 45.4 | 2.0   | 234    | 1 | 084832  | Spinocerebellar at  |
| 29 | 45.2 | 2.0   | 1505   | 1 | 055750  | Genomic clone G11H  |
| 30 | 45.2 | 2.0   | 2301   | 1 | V20445  | Human c-trk oncoge  |
| c  | 44.6 | 2.0   | 117231 | 1 | V62176  | HSV-2 strain SB5 C  |
| 32 | 44   | 2.0   | 1686   | 1 | 087587  | DNA encoding Leu    |
| 33 | 44   | 2.0   | 1995   | 1 | 014184  | N. clavipes draglin |
| 34 | 44   | 2.0   | 1995   | 1 | V23250  | N. clavipes draglin |

|      |      |     |      |   |        |                     |
|------|------|-----|------|---|--------|---------------------|
| c 35 | 44   | 2.0 | 3312 | 1 | X37085 | FMF associated pro  |
| c 36 | 43.6 | 1.9 | 154  | 1 | Q84835 | Spino cerebellar at |
| c 37 | 43.6 | 1.9 | 203  | 1 | V17226 | SCA2 gene CAG repe  |
| c 38 | 43.6 | 1.9 | 203  | 1 | V30271 | Glutamine rich reg  |
| c 39 | 43.6 | 1.9 | 913  | 1 | T07199 | Cotton fibre-speci  |
| c 40 | 43.6 | 1.9 | 913  | 1 | T13034 | Cotton fibre-speci  |
| c 41 | 43.6 | 1.9 | 913  | 1 | T30253 | Cotton fibre cell-1 |
| c 42 | 43.6 | 1.9 | 913  | 1 | T62610 | Cotton fibre speci  |
| c 43 | 43.6 | 1.9 | 913  | 1 | T70041 | Cotton fibre speci  |
| c 44 | 43.6 | 1.9 | 1884 | 1 | T13030 | Cotton fibre-speci  |
| c 45 | 43.6 | 1.9 | 1985 | 1 | T30250 | Cotton fibre clone  |

## ALIGNMENTS

|        |   |
|--------|---|
| RESULT | 1   |
| ID     | X39644  |
| AC     | X39644 standard; DNA: 2236 BP.  |
| DT     | 02-JUL-1999 (first entry)   |
| DE     | Renal cancer associated gene.   |
| KW     | Cancer associated antigen; diagnosis; research; treatment; human;<br>breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;<br>prostate cancer; SS.  |
| KW     | Homo sapiens.   |
| OS     | Homo sapiens.   |
| PN     | WO9904265-A2.   |
| PD     | 28-JAN-1999.  |
| PF     | 15-JUL-1998; UI4679.  |
| PR     | 22-JUN-1998; US-102322.   |
| PR     | 17-JUL-1997; US-896164.   |
| PR     | 10-OCT-1997; US-061599.   |
| PR     | 10-OCT-1997; US-061765.   |
| PR     | 10-OCT-1997; US-948705.   |
| PR     | 11-OCT-1997; GB-021697.   |
| PA     | (LUDWIG) LUDWIG INST CANCER RES.  |
| PI     | Chen Y, Gout I, Gure A, OHare M, Obata Y, Old LJ,<br>Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,<br>Tureci O,<br>Wefl; 99-13248/11.  |
| PT     | New isolated cancer associated nucleic acids and polypeptides -<br>isolated using sera from cancer patients, used to develop products<br>for the diagnosis, monitoring or treatment of cancers  |
| PS     | Claim 67, Page 435-436; 787pp; English.   |
| CC     | The invention relates to a method for diagnosing a disorder characterised<br>by expression of a human cancer associated antigen precursor coded for by<br>a nucleic acid molecule (NAM). The method comprises: (a) contacting a<br>biological sample isolated from a subject with an agent that specifically<br>binds to the NAM, an expression product or a fragment of an expression<br>product complexed with an HLA molecule; and (b) determining the<br>interaction between the agent and the NAM or the expression product as a<br>determination of the disorder. The products and methods can be used in<br>the diagnosis, monitoring, research, or treatment of conditions<br>characterised by the expression of various cancer associated antigens.<br>The invention provides nucleic acid sequences and encoded polypeptides<br>which are cancer associated antigen precursors expressed in human breast<br>cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and<br>lung cancer. |
| CC     | Sequence 2236 BP. 548 A; 588 C; 653 G; 447 T;   |

|                       |                 |               |           |              |
|-----------------------|-----------------|---------------|-----------|--------------|
| query Match           | 100.0%;         | Score 2236;   | DB 1;     | Length 2236; |
| Best Local Similarity | 100.0%;         | Pred. No. 0;  |           |              |
| Matches 2236;         | Conservative 0; | Mismatches 0; | Indels 0; | Gaps 0;      |

[illegible]

121 ttccgcataaagctgtagattcttcgtatctgaaatagatgcagagaaggaactatctctatgtat 180  
121 ttccgcgcataaagtgatgattttctgatttgaaatgatgcagagaaagacatctctatgtat 180  
QY 181 gtgctgcgaatgtaaccaacagacatgagctgagccgtgctgtaggaagacctgaaagctg 240  
Db 181 gtgctgcgaatgtaaccaacagacatgagctgagccgtgctgtaggaagacctgaaagctg 240  
QY 241 gtcataaataagaaacccagccgtctgctctgtttgtagtcacatcgccgctgtagcccaactg 300  
Db 241 gtcataaataagaaacccagccgtctgctctgtttgtagtcacatcgccgctgtagcccaactg 300  
QY 301 aagcacaacagctggaatataatgacgtgcaccccccgcgcgtccaggaagctgaaagagtg 360  
Db 301 aagcacaacagctggaatataatgacgtgcaccccccgcgcgtccaggaagctgaaagagtg 360  
QY 361 agctgcgaacgctgcgaaccccggaagccctcgccctgtagtgctgctgtagccgtgagttt 420  
Db 361 agctgcgaacgctgcgaaccccggaagccctcgccctgtagtgctgctgtagccgtgagttt 420  
QY 421 ggcctgtagcctcttcaatctccacatcaaaagcgctgacagcagacagcgtcgaggttc 480  
Db 421 ggcctgtagcctcttcaatctccacatcaaaagcgctgacagcagacagcgtcgaggttc 480  
QY 481 caggttagaggaacgagatcgctccgagatcaatggaatcctcctcctgtaaccaatgag 540  
Db 481 caggttagaggaacgagatcgctccgagatcaatggaatcctcctcctgtaaccaatgag 540  
QY 541 gaaggtcaataacatctcgaaaccaaagaataatgtgtccatcaataatgtagaacaatcg 600  
Db 541 gaaggtcaataacatctcgaaaccaaagaataatgtgtccatcaataatgtagaacaatcg 600  
QY 601 cgcgtatcccgtaaaaagcctctctctgtagagccctcactgtgagcatgtgtagatct 660  
Db 601 cgcgtatcccgtaaaaagcctctctctgtagagccctcactgtgagcatgtgtagatct 660  
QY 661 gctgcggaatctgtagggcgctgtaggaagcagcctgagcctccctcgtagaactcgagaaacag 720  
Db 661 gctgcggaatctgtagggcgctgtaggaagcagcctgagcctccctcgtagaactcgagaaacag 720  
QY 721 gagaagaagtgcttcaatcaagcctgtgttagagctcccgagagcctgtgctgcagcatctc 780  
Db 721 gagaagaagtgcttcaatcaagcctgtgttagagctcccgagagcctgtgctgcagcatctc 780  
QY 781 ggcacacacagaaagcctgcatcttcaatgagcctgtagaagcctgtagccctgctgtgct 840  
Db 781 ggcacacacagaaagcctgcatcttcaatgagcctgtagaagcctgtagccctgctgtgct 840  
QY 841 gaggtgaggaatgtagaataaggaacagatgtgtcgaatgcaatgtagcgtcgaatctctaac 900  
Db 841 gaggtgaggaatgtagaataaggaacagatgtgtcgaatgcaatgtagcgtcgaatctctaac 900  
QY 901 ctgtagtcaaaagagagctgtataatgtgtcgtgaataaataagccgagcctgtagccatctc 960  
Db 901 ctgtagtcaaaagagagctgtataatgtgtcgtgaataaataagccgagcctgtagccatctc 960  
QY 961 gtagctgcagcctgtagccgagagctgttcaatgtagaagcctgtaggaagcgtgtaggaagc 1020  
Db 961 gtagctgcagcctgtagccgagagctgttcaatgtagaagcctgtaggaagcgtgtaggaagc 1020  
QY 961 gtagctgcagcctgtagccgagagctgttcaatgtagaagcctgtaggaagcgtgtaggaagc 1020  
Db 961 gtagctgcagcctgtagccgagagctgttcaatgtagaagcctgtaggaagcgtgtaggaagc 1020  
QY 1021 cagcgtgtagcctgtagcagcagagagcttctcaatgtagaagcgtgtaggaagcgtgtagga 1080  
Db 1021 cagcgtgtagcctgtagcagcagagagcttctcaatgtagaagcgtgtaggaagcgtgtagga 1080  
QY 1081 aagatctctcaggaagcagcagagagctgtagcagcaggaaggaagaataatgtgcccagaag 1140  
Db 1081 aagatctctcaggaagcagcagagagctgtagcagcaggaaggaagaataatgtgcccagaag 1140  
QY 1141 gcaagcaggaagaataatgagagatccggaaggaatggaacagatgtgtaggaaggaagag 1200  
Db 1141 gcaagcaggaagaataatgagagatccggaaggaatggaacagatgtgtaggaaggaagag 1200  
QY 1201 aagtttaagaagaatgtaggaagaagaactgtaggtctcaaaaggaacagctactctgtcctaaa 1260

1201 aaatttaagaagcaatggagaaagaaagactggcgctcaaaagaaacagactctgctctaaa 1260  
QY 1261 accataacgtgctgtaggtataaccagatfaccctctcgaaagccaaagatgatacaaggaagtg 1320  
Db 1261 accataacgtgctgtaggtataaccagatfaccctctcgaaagccaaagatgatacaaggaagtg 1320  
QY 1321 gaaacctgcagcccgacagatgacccctgtagatgtaggagcaggaaggaaggaagcag 1380  
Db 1321 gaaacctgcagcccgacagatgacccctgtagatgtaggagcaggaaggaaggaagcag 1380  
QY 1381 gattccggaataatgaaagaagccttgaaacccctactatgatagtctcccaagagagatc 1440  
Db 1381 gattccggaataatgaaagaagccttgaaacccctactatgatagtctcccaagagagatc 1440  
QY 1441 atgaggaaagatgtccggtctccacacatcaaaagaaggaagatccttgaagcctgagccctg 1500  
Db 1441 atgaggaaagatgtccggtctccacacatcaaaagaaggaagatccttgaagcctgagccctg 1500  
QY 1501 gaaagcgtgtagactcccccacttgaggaaagtgtagctgtctgtgtagaagcagggga 1560  
Db 1501 gaaagcgtgtagactcccccacttgaggaaagtgtagctgtctgtgtagaagcagggga 1560  
QY 1561 gctgctgagcagcaatgtagcattgtgaaagggagcagatcaatggaatcaacagcagag 1620  
Db 1561 gctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1620  
QY 1621 atgtgtacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
Db 1621 atgtgtacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
QY 1681 ggcgaggaactgtagcagcctgtgtagcgtctgtagcccccagaaaggaatgtagaagtag 1740  
Db 1681 ggcgaggaactgtagcagcctgtgtagcgtctgtagcccccagaaaggaatgtagaagtag 1740  
QY 1741 ctgagccttctgctgtagaagtcgaaggggaaccaaattcaacgctgttaggaacagtgag 1800  
Db 1741 ctgagccttctgctgtagaagtcgaaggggaaccaaattcaacgctgttaggaacagtgag 1800  
QY 1801 ctccggccccacactctgtagaacaagaagcctcgaaacagccttgaaggaagcagcagtaga 1860  
Db 1801 ctccggccccacactctgtagaacaagaagcctcgaaacagccttgaaggaagcagcagtaga 1860  
QY 1861 caccacagatgtagcactctgtagaagcctgtagaagccttcaacccagaaatcacaatcccttg 1920  
Db 1861 caccacagatgtagcactctgtagaagccttcaacccagaaatcacaatcccttg 1920  
QY 1921 ggcctgaaacagggccagataaagaagaagcctgagccactttttgaaagcgaatgtgga 1980  
Db 1921 ggcctgaaacagggccagataaagaagaagcctgagccactttttgaaagcgaatgtgga 1980  
QY 1981 ggaagaaggaagcagcagcagccttgggaagaagatcctaagaagatcagaactcaatctct 2040  
Db 1981 ggaagaaggaagcagcagcagccttgggaagaagatcctaagaagatcagaactcaatctct 2040  
QY 2041 cctctgagccagtgaaatgtgtctctccagccttggggagactccttcttgaaaccttaa 2100  
Db 2041 cctctgagccagtgaaatgtgtctctccagccttggggagactccttcttgaaaccttaa 2100  
QY 2101 taagaccccaatgtaggtctctctctccatccctctctctgagcctgagcttcaattgc 2160  
Db 2101 taagaccccaatgtaggtctctctctccatccctctctctgagcctgagcttcaattgc 2160  
QY 2161 tgcacagatgtgtactcctcaaaccttactctgagcctatgtatgtatgtatgtatgtatgt 2220  
Db 2161 tgcacagatgtgtactcctcaaaccttactctgagcctatgtatgtatgtatgtatgtatgt 2220  
QY 2221 ttccagcttaaaaaa 2236  
Db 2221 ttccagcttaaaaaa 2236

RESULT 2

| ID                    | Query Match   | Best Local Similarity | Matches 2236: | Conservative | Mismatches | Indels | Gaps |
|-----------------------|---|-----------------------|---------------|--------------|------------|--------|------|
| X40057                | standard; DNA.  | 2236 BP.              |               |              |            |        |      |
| AC                    | X40057;   |                       |               |              |            |        |      |
| DR                    | 02-JUL-1999 (first entry)   |                       |               |              |            |        |      |
| DE                    | Colon cancer associated gene.   |                       |               |              |            |        |      |
| KM                    | Cancer associated antigen; diagnosis; research; treatment; human;         |                       |               |              |            |        |      |
| KW                    | breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;   |                       |               |              |            |        |      |
| KM                    | prostate cancer; ss.  |                       |               |              |            |        |      |
| OS                    | Homo sapiens.   |                       |               |              |            |        |      |
| PN                    | WO904265-A2.  |                       |               |              |            |        |      |
| PD                    | 28-JAN-1999.  |                       |               |              |            |        |      |
| PF                    | 15-JUL-1998; U14679.  |                       |               |              |            |        |      |
| PR                    | 22-JUN-1998; US-102332.   |                       |               |              |            |        |      |
| PR                    | 17-JUL-1997; US-896164.   |                       |               |              |            |        |      |
| PR                    | 10-OCT-1997; US-061599.   |                       |               |              |            |        |      |
| PR                    | 10-OCT-1997; US-061765.   |                       |               |              |            |        |      |
| PR                    | 10-OCT-1997; US-948705.   |                       |               |              |            |        |      |
| PR                    | 11-OCT-1997; GB-021697.   |                       |               |              |            |        |      |
| PA                    | (LUDWIG) LUDWIG INST CANCER RES.  |                       |               |              |            |        |      |
| PI                    | Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,                         |                       |               |              |            |        |      |
| PI                    | Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,                          |                       |               |              |            |        |      |
| PI                    | Tureci O;   |                       |               |              |            |        |      |
| DR                    | WPI: 99-132448/11.  |                       |               |              |            |        |      |
| PT                    | New isolated cancer associated nucleic acids and polypeptides -           |                       |               |              |            |        |      |
| PT                    | isolated using sera from cancer patients, used to develop products        |                       |               |              |            |        |      |
| PT                    | for the diagnosis, monitoring or treatment of cancers                     |                       |               |              |            |        |      |
| PS                    | Claim 67; Page 655-656; 787pp; English.                                   |                       |               |              |            |        |      |
| CC                    | The invention relates to a method for diagnosing a disorder characterised |                       |               |              |            |        |      |
| CC                    | by expression of a human cancer associated antigen precursor coded for by |                       |               |              |            |        |      |
| CC                    | a nucleic acid molecule (NAM). The method comprises: (a) contacting a     |                       |               |              |            |        |      |
| CC                    | biological sample isolated from a subject with an agent that specifically |                       |               |              |            |        |      |
| CC                    | binds to the NAM, an expression product or a fragment of an expression    |                       |               |              |            |        |      |
| CC                    | product complexed with an HLA molecule; and (b) determining the           |                       |               |              |            |        |      |
| CC                    | interaction between the agent and the NAM or the expression product as a  |                       |               |              |            |        |      |
| CC                    | determination of the disorder. The products and methods can be used in    |                       |               |              |            |        |      |
| CC                    | the diagnosis, monitoring, research, or treatment of conditions           |                       |               |              |            |        |      |
| CC                    | characterised by the expression of various cancer associated antigens.    |                       |               |              |            |        |      |
| CC                    | The invention provides nucleic acid sequences and encoded polypeptides    |                       |               |              |            |        |      |
| CC                    | which are cancer associated antigen precursors expressed in human breast  |                       |               |              |            |        |      |
| CC                    | cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and   |                       |               |              |            |        |      |
| CC                    | lung cancer.  |                       |               |              |            |        |      |
| SQ                    | Sequence  | 2236 BP;              | 548 A;        | 588 C;       | 653 G;     | 447 T; |      |
| Query Match           | 100.0%;   | Score 2236;           | DB 1;         | Length 2236; |            |        |      |
| Best Local Similarity | 100.0%;   | Pred. No. 0;          |               |              |            |        |      |
| Matches 2236:         | Conservative  | 0;                    | Mismatches    | 0;           | Indels     | 0;     | Gaps |
| 0Y                    | 1 cctggcccggtcgcggtcgcggtcttccagctccttgccgagccggccgagccgaagaac            | 60                    |               |              |            |        |      |
| DB                    | 1 cctggcccggtcgcggtcgcggtcttccagctccttgccgagccggccgagccgaagaac            | 60                    |               |              |            |        |      |
| 0Y                    | 61 gggctgctgaagaagcagctgagctggccagccatgagccgaagaagtggcccgagaa             | 120                   |               |              |            |        |      |
| DB                    | 61 gggctgctgaagaagcagctgagctggccagccatgagccgaagaagtggcccgagaa             | 120                   |               |              |            |        |      |
| 0Y                    | 121 ttccggcaataagtgatatttcttgattgtaaaatgtagtcgagaagaagacatctatagt         | 180                   |               |              |            |        |      |
| DB                    | 121 ttccggcaataagtgatatttcttgattgtaaaatgtagtcgagaagaagacatctatagt         | 180                   |               |              |            |        |      |
| 0Y                    | 181 gttctgtcgaaatgtaccaccagacacatgtagcgtggccgtctcgttggagaaactgaagctg      | 240                   |               |              |            |        |      |
| DB                    | 181 gttctgtcgaaatgtaccaccagacacatgtagcgtggccgtctcgttggagaaactgaagctg      | 240                   |               |              |            |        |      |
| 0Y                    | 241 gttcatcaatgaaccagccgctgcctctgtttgatgcatctcggcgtgatcccaatg             | 300                   |               |              |            |        |      |
| DB                    | 241 gttcatcaatgaaccagccgctgcctctgtttgatgcatctcggcgtgatcccaatg             | 300                   |               |              |            |        |      |
| 0Y                    | 301 aagccacgaagtgaatatatatactgagcccccggcgctccagaagaagtgtgaagagtg          | 360                   |               |              |            |        |      |
| DB                    | 301 aagccacgaagtgaatatatatactgagcccccggcgctccagaagaagtgtgaagagtg          | 360                   |               |              |            |        |      |
| 0Y                    | 361 cgtctggagccgtctgcaccccggaagcctcggccttgaagtgtcgttggctgagattc           | 420                   |               |              |            |        |      |

Db 1441 AAGGGAAGAGATGTCGGCTCTACGACATCAGAAGAGAGGATCCTTGAAGCTGGCCCTG 1500  
Qy 1501 gaagcggtgtgagactcccccattgggaagtggtcgttcgtgtgtaagcg99ga 1560  
Db 1501 |||||  
Db 1501 GAAGGGGTGTGACTCCCATTTGGGAAGTGTGTTCTCTGTGTATGACGGGGA 1560  
Qy 1561 gctgtcgaagcagatgtgtgcatctgtgaaaggagagatcatggaatcaacg99caag 1620  
Db 1561 |||||  
Db 1561 GCTGCTGAGCGGCGATGTGGCATTTGTGAAGAGGAGAGATCATGCAATCAACGGCAAG 1620  
Qy 1621 attgtgacagactaacacccctgtgtgagctgtgaagctgtccctgtgaagaagcgtggaatcag 1680  
Db 1621 |||||  
Db 1621 ATTGTATACAGACTACACCTGTGCTGAGCTGAGCTGCTGTGAGAGGGCTTGAATCAG 1680  
Qy 1681 ggcggagactgtgacacacctgtgtgtccgtgtcccccgaaggatgacgtgag 1740  
Db 1681 |||||  
Db 1681 GCGGGGACTGATTCACCTTGTGTGTTGCCGTGTGCCCCCAAGAGATGTGACATGAG 1740  
Qy 1741 ctgacctctgtgtgaagtcacaaaggagagaaattcaagcgttaagaaacagttag 1800  
Db 1741 |||||  
Db 1741 CTGACCTTCTGCTGAAGTCCAAAGGGGAACCAATTCACGCTTGAAGAACAGTAG 1800  
Qy 1801 ctccggccccacactgtgtgaacaaagcctcgtgaagcctgtgagagagccacatgaca 1860  
Db 1801 |||||  
Db 1801 CTCCGGCCCCACTCTGTGAACACAAAGCCTCGGACACAGCTTGAAGAGGCCACATGACA 1860  
Qy 1861 caacacagatgtgacatcttggagacctgaatctatacccccagagatctcaaacctcttg 1920  
Db 1861 |||||  
Db 1861 CACACACAGATGGATCTTGTGGACCTGATCTTACACCAAGAAATCTCAAACTCCCTTGTG 1920  
Qy 1921 gccctgaacagagggccagataaagaaacagctcgggacactttttgaagggcaatgtgaa 1980  
Db 1921 |||||  
Db 1921 GGCCTGAACAGGGGCGAGATAAGAAAGACGCTGGGCGACATTTTGAAGGCCAATGTGGA 1980  
Qy 1981 ggaagagggagcgccagccgcttgggagaaagatctcaagatccaaactctatctctt 2040  
Db 1981 |||||  
Db 1981 GGAAGAGGAGCACCGCGTTGGAGAAAGATCTCAAGATCCACACTCTCTTCTTCTT 2040  
Qy 2041 cctctggccagatgaatttgtctctccagacttggggagactccttccttgaacctta 2100  
Db 2041 |||||  
Db 2041 CCTCTGGCCAGTGAATTTGGTCTCTCCAGACTTTTGGGAGCTCTTCTTGAACCTTAA 2100  
Qy 2101 taagaccacactgtgagctctctctctcatccctctcctcgtccctgtcttaattg 2160  
Db 2101 |||||  
Db 2101 TAAAGACCCACATGAGATCTCTCTCTCATCCCTCTCTGCTGCCCTGCTTAATTTG 2160  
Qy 2161 tgcagagatgtgacatcccaacttaactgtgagctcataataataaacaagattat 2220  
Db 2161 |||||  
Db 2161 TGCAGAGATTGTCTACTCCAAACCTTACTGTAGCTATTAAATTAACAGATTATT 2220  
Qy 2221 ttccagcttaaaaaa 2236  
Db 2221 |||||  
Db 2221 TTCCAGCTTAAAAAAA 2236

PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohare M, Obara Y, Old LJ,  
PI Pfeundschn M, Sahin U, Scanlan MJ, Stockert E,  
PI Tureci O;  
DR WPI; 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT Isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67, Page 663-664; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2289 BP; 567 A; 596 C; 668 G; 458 T;  
Query Match 97.2%; Score 2173; DB 1; Length 2289;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 2236; Conservative 0; Mismatches 0; Indels 53; Gaps 1;  
Qy 1 cctggccgctgcggtgcggtccttcagctcctgtgcaagcgagccacgaagaaac 60  
Db 1 |||||  
Db 1 CCTGGCCCGGTGCCGGTCCGGCTCTTCCAGCTCCTGCAAGCGGGGCAACCAAGAAAC 60  
Qy 61 gggctgtcaacagcagcagctgacactggccacgacatgagacgaaagtgtgcccagaa 120  
Db 61 |||||  
Db 61 GGGCTGTCAACAGCAGCAGCTGAGACTGGCCAGCATGAGACCGAAAGTGGCCCGAAGA 120  
Qy 121 ttccggcataagtgatatttctgattgaaatgagcagagagaagacatctctatgt 180  
Db 121 |||||  
Db 121 TTCCGGCATTAAGTGGATTTTCTGATTGAAAAGATGCAAGAGACATATCTTATGAT 180  
Qy 181 gtgtcgtgaatgtaccacacagacatgtgacgtgtgcgtgtgtgagaaactgtgacgt 240  
Db 181 |||||  
Db 181 GTCTCTCGAATGTACACACAGACATGACATGACGCTGCTGTGGAGACCTGAAAGCTG 240  
Qy 241 gtcatcaatgaaccacagcgcgtctgtcctgtgtgattgcatcattgcgcgtgtatccactg 300  
Db 241 |||||  
Db 241 GTCATCAATGAACCCAGCGGTGCTGTGTTGATGCCATTGCGCGCTGATCCCATG 300  
Qy 301 aagcaccaagtggaatatgatacagctgaccccccgctccaggaagctgaaagaggtg 360  
Db 301 |||||  
Db 301 AAGCACCAAGTGAATATATATACGTGACCCCGCGCTCCAGGAAGCTGAAGAGGTG 360  
Qy 361 cgtctgagacgtctgcaccccgaaagcctcgcgtgtgagtgctgtgtgtgctgagattc 420  
Db 361 |||||  
Db 361 CGTCTGACCGTGTGTGACCCCGAAGGCTTCGCTAAGTGTGTGTGTGTGTGTGTGTGT 420  
Qy 421 ggcgtgtgagccttcaatcccaactcaatcaaaagcgttcagagcagcgtcgtggttc 480  
Db 421 |||||  
Db 421 GGCCTGGGCGCTCTTCAATCTCCACCTCATCAAAAGGGGTGAGGACAGACGCGTGGCTC 480  
Qy 481 caggtgagggagagatgctgcgagatcatgatatcatctccctcctgtatcccatgag 540  
Db 481 |||||  
Db 481 CAGGTAGGGGAGAGATCTCGGATCAATGATATTCATCTCCCTGATCCCATGAG 540  
Qy 541 gaggtcatcaacctcatctgaaccaagaagaactgttcatcatcaagtgtgaacatcgtgc 600  
Db 541 |||||  
Db 541 GAGGTATATACCTCTTGTGAACCAAGAAACTGTGTCCATCAAGTGAACATCATCGGC 600  
Qy 601 ctgatacccgtaaaaagcctctcgtatgagccctcaacttggcagatgtgtgatcagtt 660  
Db 601 |||||

Db 601 CTGATCCCGTGAAGAGCTCTCTGATGAGCCCTTCACTTGCGAGTATGTGATCAGTTT 660  
QY 661 ggtgtggaatcttgaggcgtgagagcagccttggtctcccttggaatctgggaaacaag 720  
Db 661 gttctcgaatcttgaggcgtgagagcagccttggtctcccttggaatctgggaaacaag 720  
QY 721 gaaagaaggtctctcaacccctgttaggtctccgagccttggtctgcaacattccagc 780  
Db 721 GAAAGAAGGTCTTCAATCAACCTGTGAGGCTCCCGAGGCTTGGCTGACATTTCCAGC 780  
QY 781 ggcacatccagaagccttgcatcttatacagcaltgtgaacctggtctccctgtctgt 840  
Db 781 GGCACATCCAGAAGCCTGGCATCTTTATCAGCATGTGAACCTGGCTCCCTGTGCTGCT 840  
QY 841 gagggtggtatgtgagatagggagcagatctgtcgaagtcaatggcgtcgaactctaac 900  
Db 841 GAGGTGGGATGTGAGATAGGGAGCAGATGTGCAATGGCTGCGATCTCTCTAAC 900  
QY 901 ctggtacacaagaagcctgttaaatgtctgaataataagccgagccttgaccatctccatt 960  
Db 901 CTGTGATCACAAGAGAGCTGTAAATGTGCTGAAAAATAGCCGACGCTTGACCATCTCCATT 960  
QY 961 gtagctcagctctgagcggagcgtgtctatagaacagccggagcgtctgagagcgcg 1020  
Db 961 GTAGCTGACAGCTGGCCGGAGCTGTTTCATACAGACCGGAGGCGCTGGCAGAGGCGCG 1020  
QY 1021 cagcgttgaactcagcggcagagagccttctcaatgcaagaacgctgagcgtatgagttccaa 1080  
Db 1021 CAGCGTGAAGCTGACGCGGCGAGAGCTTCATCAGAACGCGCTGGGATGGAGTCCAAC 1080  
QY 1081 aagatctctcagaagcagcagagatctgagcggcgaagaagaagaatctgcagaag 1140  
Db 1081 AAGATCTCTCAGAGAGAGCAGGAGATGAGAGCGGCAAAAGGAAATTTGCCCAAG 1140  
QY 1141 gcaagcagaagaataagagataccggaagagatggaacagatctgtagagagagagag 1200  
Db 1141 GCAGCAGAGGAAATGAGAGATACCGGAAGAGATGGAACAGATTGTAAGAGAGAAAG 1200  
QY 1201 aagtttaagaagaatgaggaagaagacttggtctcaagaagaacagctactcttgcttaaa 1260  
Db 1201 AAGTTTAAGAAGCAATGGGAAAGACTGGGCTCAAAAGGAAACGATCACTTGGCCCTAAA 1260  
QY 1261 accatctgctgaggtatgacccagtaacctctgcgaagcgaagtatgtagcagagtg 1320  
Db 1261 ACCATCTACTGCTAGGTATACCCAGTACCCCTTCCGACGCAATGTGATCGAGAGTG 1320  
QY 1321 gaacctgagctcagcggcgcagatgaactgtgagagcagcggagagagcagcag 1380  
Db 1321 GAACCTGAGCTCGAGCCGCGAGATGACCTGATGAGAGCGACGAGGAGAGAGAG 1380  
QY 1381 -----gattcc 1387  
Db 1381 CCACAGAGAGATGTTGAAGAGATGTGTTTATCAAGACAGCATTCACAGATTTCC 1440  
QY 1388 gpaataatgagagagccttgacccctactctatgttcaaccagagcagatgagga 1447  
Db 1441 GGAATAATGAGAGAGCTTTGACCCCTACTCTATGTTTACCCAGAGCAATATATGGGGA 1500  
QY 1448 agagatgtccggtctctacgcatcaagaagaagagatccttagaaccttgccctggaagcg 1507  
Db 1501 AGGATGTCCGGCTCTACGATCAAGAAGAGGATCTTGAACCTGCGCTCGGAAGGCG 1560  
QY 1508 gttgtggaatcccccattgggaaggtggtcttctcgtgtatgagcgggagcgtgctg 1567  
Db 1561 GTGTGACTCCCTCATTTGGGAAGGTGTCTTTCTGCTGTGTGAGCGGGAGCTGCTG 1620  
QY 1568 agggagatgtgagcattgtgaagaggagacagatcatgtgaatcaacaggaagaattgtga 1627  
Db 1621 AGCGGATGTGCTGATTTGTAAGAGGAGAGATCAATGGAATCAACGCAATATGTGTA 1680  
QY 1628 cagactacacccctggtctgagcgtctgacgtctgcccctgagaagccttggaatcagggcg 1687  
Db 1681 CAGACTACACCCCTGTGGCTGAGGCTGACGCTCCCTGGCAGAGGCTTGAATCAGGCGGGG 1740

QY 1688 actgtagcagccttggtgtgctgcccgtctgcccccaaaagagatagagatgagcct 1747  
Db 1741 ACTGATGTGACCTTGTGTTGGCTGCTGCCCCCAAGAGATGAGATGAGTACCTGACCT 1800  
QY 1748 tcttctgtaagtcacaaagggagaaaccaatctacaggttaggaagaagtgaagcctggc 1807  
Db 1801 TCTTGTGAGTCCAAAGGGAGAAACCAATTTCAAGCGGTTAGGAACAGTGAAGTCCGGC 1860  
QY 1808 cccaactgtgaaacaaaagcctcggagcagccttgagagagggccacatgacacacaa 1867  
Db 1861 CCCACCTCTGTAACACAAAGCCTCGAGACAGCCTTGAGAGAGGCAATGACACACACA 1920  
QY 1868 gatgtagcctcttgagacgttaactatcaccacgaagatctcaactccttggccctga 1927  
Db 1921 GATGTGCACTCTTGGAGACTTAATCTATACACAGGAATCTCAAACTCTTGGCCCTGA 1980  
QY 1928 accagggccagataaagaagcagcctcgggacactttttgaaagccaaatgtgagagaaag 1987  
Db 1981 ACCAGGGCCAGATRAGAACAGCTCGGCGCATTTTGAAGGCCAATGTGAGAGAAAG 2040  
QY 1988 gaggcagcagcgttttgagagagatctcaaggatccagactctcattccttctctg 2047  
Db 2041 GAGCAGCCACGCCCTTGGAGAGATCTCAAGGATCCAGACTGTCAATTCCTTCTCTGG 2100  
QY 2048 cccaatgtaattgtctctccagccttgaggagcctcctccttgaaacctataaagac 2107  
Db 2101 CCCAGTAAATTTGGTCTCTCCAGCTTGGGGAGACTCTTCACTTGAACCTTAATTAAGCC 2160  
QY 2108 ccaatgagctctctctccatccctcctcctcctcctcctcctcctcctcctcctcct 2167  
Db 2161 CCACTGAGAGTCTCTCTCTCTCCATCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220  
QY 2168 attgtcactcacaaccttactctgagctcatataaataaacaagattatltccagc 2227  
Db 2221 ATTGTCACTCAAAACCTTACTGTGAGCTCATTTATAAATAACAGATTATTTTCCAGC 2280  
QY 2228 ttaaaaaa 2236  
Db 2281 TTAATAAAA 2289

RESULT 4  
X40058  
ID X40058 standard; DNA; 2162 BP.  
AC X40058;  
DT 02-JUL-1999 (first entry)  
DE Colon cancer associated gene.  
KW Cancer associated antigen; diagnosis; research; treatment; human;  
breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
prostate cancer; ss.  
OS Homo sapiens.  
PN M09904265-A2.  
PD 28-JAN-1999.  
PF 15-JUL-1998; U14679.  
PR 22-JUN-1998; US-102322.  
PR 17-JUL-1997; US-896164.  
PR 10-OCT-1997; US-061599.  
PR 10-OCT-1997; US-061765.  
PR 10-OCT-1997; US-948705.  
PR 11-OCT-1997; GB-021697.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Chen Y, Gout I, Gure A, Ohare M, Ohta Y, Old LJ,  
PI Pfrendrich M, Sahin U, Scanlan MD, Stockert E,  
PI Tureci O;  
DR MPI: 99-132448/11.  
PT New isolated cancer associated nucleic acids and polypeptides -  
PT isolated using sera from cancer patients, used to develop products  
PT for the diagnosis, monitoring or treatment of cancers  
PS Claim 67; Page 658; 787pp; English.  
CC The invention relates to a method for diagnosing a disorder characterised  
CC by expression of a human cancer associated antigen precursor coded for by  
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a

CC biological sample isolated from a subject with an agent that specifically  
CC binds to the NAM, an expression product or a fragment of an expression  
CC product complexed with an HLA molecule; and (b) determining the  
CC interaction between the agent and the NAM or the expression product as a  
CC determination of the disorder. The products and methods can be used in  
CC the diagnosis, monitoring, research, or treatment of conditions  
CC characterised by the expression of various cancer associated antigens.  
CC The invention provides nucleic acid sequences and encoded polypeptides  
CC which are cancer associated antigen precursors expressed in human breast  
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
CC lung cancer.  
SQ Sequence 2162 BP; 528 A; 573 C; 622 G; 439 T;

Query Match 92.9%; Score 2078; DB 1; Length 2162;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 2162; Conservative 0; Mismatches 0; Indels 74; Gaps 1;

QY 1 ccttgccggtcgcggtcgcggtccttcacagctcctgcaagcgggcaaccgaaagaaac 60  
DB 1 cctggcccggtgcggtcgcggtccttcacagctcctgcaagcgggcaaccgaaagaaac 60  
QY 61 gggctgtcaacagcagcagctggaacctggcccaagcctgaagcgaagtgcccgagaa 120  
DB 61 gggctgtcaacagcagcagctggaacctggcccaagcctgaagcgaagtgcccgagaa 120  
QY 121 ttccgcgataaagtgatattcttcgatgaaatgatgcagagagaagactatctctatgat 180  
DB 121 ttccgcgataaagtgatattcttcgatgaaatgatgcagagagaagactatctctatgat 180  
QY 181 gtcgtcgagatgtaccacacagacatgtgacgtgacgtgctgctggtggaagacctgaacctg 240  
DB 181 gtcgtcgagatgtaccacacagacatgtgacgtgacgtgctgctggtggaagacctgaacctg 240  
QY 241 gtcataaagaaccccgccgtctgctctgtttgatgcattggccgctgatacccaagc 300  
DB 241 gtcataaagaaccccgccgtctgctctgtttgatgcattggccgctgatacccaagc 300  
QY 301 aagcacaagtgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat 360  
DB 301 aagcacaagtgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgatgat 360  
QY 361 cgtctgagacgtctgcaccccgaaagcctcgagctgagtgtgcgtgtgtgtgtgtgtgtgt 420  
DB 361 cgtctgagacgtctgcaccccgaaagcctcgagctgagtgtgcgtgtgtgtgtgtgtgt 420  
QY 421 ggcctgtgagctcttatctcccaacctcatcaaaaggggtcagcagcagcgtcgcgctc 480  
DB 421 ggcctgtgagctcttatctcccaacctcatcaaaaggggtcagcagcagcgtcgcgctc 480  
QY 481 caggtagagagagagatcgtccggatcaatgtgatgtatctcctcctgtacccaatgag 540  
DB 481 caggtagagagagagatcgtccggatcaatgtgatgtatctcctcctgtacccaatgag 540  
QY 541 gaggatcatcaacctcatcgaaccaaagaactgtgtccatcaaaagtgaagacacatcgcgc 600  
DB 541 gaggatcatcaacctcatcgaaccaaagaactgtgtccatcaaaagtgaagacacatcgcgc 600  
QY 601 ctgattcccggtgaaagcctctcgtatgagccctcactgtgcaagtatgtatcaagttt 660  
DB 601 ctgattcccggtgaaagcctctcgtatgagccctcactgtgcaagtatgtatcaagttt 660  
QY 661 gttcggaaatctgagggcgtgcaagcagcctgagcctcctcctgaagatcggaagaaacag 720  
DB 661 gttcggaaatctgagggcgtgcaagcagcctgagcctcctcctgaagatcggaagaaacag 720  
QY 721 gagaagaagtgctcagcctgtgtaggtcccgagagccttggtgtcagcatcttcacagc 780  
DB 721 gagaagaagtgctcagcctgtgtaggtcccgagagccttggtgtcagcatcttcacagc 780  
QY 781 ggcacacatcagaagcctgtgcatcttatacagcagatgtgaagcgtggtcctgtcgtctgt 840  
DB 781 ggcacacatcagaagcctgtgcatcttatacagcagatgtgaagcgtggtcctgtcgtctgt 840

QY 841 gagggtgagattgagataagggagccagatgtcgaagtcgaatgacgtgcactctctaac 900  
DB 841 gagggtgagattgagataagggagccagatgtcgaagtcgaatgacgtgcactctctaac 900  
QY 901 ctgatatcaagaagagctgttaaatgtgtctgaaataatagccgaagcctgacatctcat 960  
DB 901 ctgatatcaagaagagctgttaaatgtgtctgaaataatagccgaagcctgacatctcat 960  
QY 961 gtagctgcagctgcccggagagctgtctcatgacagacccggagggctgacagagggcg 1020  
DB 961 gtagctgcagctgcccggagagctgtctcatgacagacccggagggctgacagagggcg 1020  
QY 1021 caagctgagctgcagcagcagcagcagccttcatactgcaagcagcgtgctgagatccaac 1080  
DB 1021 caagctgagctgcagcagcagcagcagccttcatactgcaagcagcgtgctgagatccaac 1080  
QY 1081 aagatcctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1140  
DB 1081 aagatcctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1140  
QY 1141 gcaagcagaagaaatgagagatgacccggaagagatgagagagatgagagagagagag 1200  
DB 1141 gcaagcagaagaaatgagagatgacccggaagagatgagagagatgagagagagagag 1200  
QY 1201 aagtttaagaagaatgag 1260  
DB 1201 aagtttaagaagaatgag 1260  
QY 1261 accatcaactgtctgaggttacaccacagtaaccttcgcaagccaaaglaagatcaagga 1320  
DB 1261 accatcaactgtctgaggttacaccacagtaaccttcgcaagccaaaglaagatcaagga 1320  
QY 1321 gaaacctgagctgagagcccgagatgacccgtgagtgagagcaagagagagagagagag 1380  
DB 1321 gaaacctgagctgagagcccgagatgacccgtgagtgagagcaagagagagagagagag 1380  
QY 1381 gattccggaatatgag 1440  
DB 1381 gattccggaatatgag 1440  
QY 1441 atggggaagagatgtccggtcctcaagcagcaagagagagagagagagagagagagag 1500  
DB 1441 atggggaagagatgtccggtcctcaagcagcaagagagagagagagagagagagagag 1500  
QY 1501 gaaagcgtgtgagacccccaattgggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560  
DB 1501 gaaagcgtgtgagacccccaattgggaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1560  
QY 1561 gctgtcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1620  
DB 1561 gctgtcgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1620  
QY 1621 attgtgacagactaacacccctgagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
DB 1621 attgtgacagactaacacccctgagcagcagcagcagcagcagcagcagcagcagcagcag 1680  
QY 1681 ggcgggagactgagatgagacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740  
DB 1681 ggcgggagactgagatgagacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1740  
QY 1741 ctgacccctctgtgagagcgaag 1800  
DB 1741 ctgacccctctgtgagagcgaag 1800  
QY 1801 ctcgagcccaacctcgttgagacaaagcctggaacacagccttgagagagagagagagag 1860  
DB 1801 ctcgagcccaacctcgttgagacaaagcctggaacacagccttgagagagagagagagag 1860  
QY 1861 cagacacagatgacatccttgag 1920  
DB 1861 cagacacagatgacatccttgag 1920  
QY 1921 ggcacacagatgacatccttgag 1980  
DB 1921 ggcacacagatgacatccttgag 1980



|                       |  |  |              |
|-----------------------|--|--|--------------|
| QY                    | 1921   | gccctgaaccacggcgccagataaagaaacgctcggggccacttlttgaagggccaatgtga | 1980         |
| Db                    | 1847   | gocctgaacacgagggccacataagaacagcttcggggccacttlttgaagggccaatgtga | 1906         |
| QY                    | 1981   | ggaagggagacgcacgcgcttgggaagaagatctcaagatccagactctatccctt       | 2040         |
| Db                    | 1907   | ggaaaggagacgacgacgcttgggaagaagatctcaagatccagactctatccctt       | 1966         |
| QY                    | 2041   | cctctgcccagatgaatttgcctctccagatttggggagacttcccttgaacctaa       | 2100         |
| Db                    | 1967   | cctctgcccagatgaatttgcctctccagatttggggagacttcccttgaacctaa       | 2026         |
| QY                    | 2101   | taagaccccaactgtagctctctctctccatccctctctctgcccctgtcctaatgc      | 2160         |
| Db                    | 2027   | taagaccccaactgtagctctctctctccatccctctctctgcccctgtcctaatgc      | 2086         |
| QY                    | 2161   | tgcagagatttgcattcccaaaccttacctgtgactgcattcaataaataaagatttat    | 2220         |
| Db                    | 2087   | tgcagagatttgcattcccaaaccttacctgtgactgcattcaataaataaagatttat    | 2146         |
| QY                    | 2221   | ttccagacttaaaaaa   | 2236         |
| Db                    | 2147   | ttccagacttaaaaaa   | 2162         |
| RESULT                | 5  |  |              |
| ID                    | X40062   |  |              |
| AC                    | X40062:  |  |              |
| DT                    | 02-JUL-1999  | (first entry)  |              |
| DE                    | Colon cancer associated gene.  |  |              |
| KW                    | Cancer associated antigen; diagnosis; research; treatment; human; breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer; prostate cancer; ss. |  |              |
| OS                    | Homo sapiens.  |  |              |
| PN                    | MO9904265-A2.  |  |              |
| PD                    | 28-JAN-1999.   |  |              |
| PF                    | 15-JUL-1998;   | U14679.  |              |
| PR                    | 22-JUN-1998;   | U1-102322.   |              |
| PR                    | 17-JUL-1997;   | US-896164.   |              |
| PR                    | 10-OCT-1997;   | US-061599.   |              |
| PR                    | 10-OCT-1997;   | US-061765.   |              |
| PR                    | 10-OCT-1997;   | US-948705.   |              |
| PR                    | 11-OCT-1997;   | GB-021697.   |              |
| PA                    | (LUDWIG) LUDWIG INST CANCER RES.   |  |              |
| PI                    | Chen Y, Gout I, Gure A, Ohare M, Ohata Y, Old LJ, Pfeundschnuh M, Sahlin U, Scanlan MJ, Stockert E, Tureci O;  |  |              |
| PI                    | WPI: 99-132448/11.   |  |              |
| PT                    | New isolated cancer associated nucleic acids and polypeptides -  |  |              |
| PT                    | isolated using sera from cancer patients, used to develop products   |  |              |
| PT                    | for the diagnosis, monitoring or treatment of cancers  |  |              |
| PS                    | Claim 67, Page 665-666; 787bp; English.  |  |              |
| CC                    | The invention relates to a method for diagnosing a disorder characterised  |  |              |
| CC                    | by expression of a human cancer associated antigen precursor coded for by  |  |              |
| CC                    | a nucleic acid molecule (NAM). The method comprises: (a) contacting a  |  |              |
| CC                    | biological sample isolated from a subject with an agent that specifically  |  |              |
| CC                    | binds to the NAM, an expression product or a fragment of an expression   |  |              |
| CC                    | product complexed with an HLA molecule; and (b) determining the  |  |              |
| CC                    | interaction between the agent and the NAM or the expression product as a   |  |              |
| CC                    | determination of the disorder. The products and methods can be used in   |  |              |
| CC                    | the diagnosis, monitoring, research, or treatment of conditions  |  |              |
| CC                    | characterised by the expression of various cancer associated antigens.   |  |              |
| CC                    | The invention provides nucleic acid sequences and encoded polypeptides   |  |              |
| CC                    | which are cancer associated antigen precursors expressed in human breast   |  |              |
| CC                    | cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  |  |              |
| CC                    | lung cancer.   |  |              |
| SC                    | Sequence   | 2409 BP;   | 595 A;       |
|                       |  | 638 C;   | 692 G;       |
|                       |  |  | 484 T;       |
| Query Match           | 91.8%;   | Score 2053;  | DB 1;        |
| Best Local Similarity | 92.8%;   | Pred. No. 0;   | Length 2409; |
| Matches 2236;         | Conservative   | 0;   | Mismatches   |
|                       |  | 0;   | Indels 173;  |
|                       |  |  | Gaps 1;      |

[illegible]

|    |      |                             |                             |      |
|----|------|-----------------------------|-----------------------------|------|
| OY | 1081 | aagatccctccagagcgccgagagatc | tgagcgccgaaggaagaatttgcgaag | 1140 |
| Db | 1081 | AAGATCCCTCCAGAGCGAGGAGATG   | AGCAGGAAGAAATTTGCCAGAG      | 1140 |
| OY | 1141 | gcagcagaggaatactgagataccgga | aggaatgaaacagattgttagaggaag | 1200 |
| Db | 1141 | GCAGCAGGAGAAATGAGAGATTAC    | CGGAGAGATGGAACAGTTGTAGAG    | 1200 |
| OY | 1201 | aagttaaagcaatgaggagaagact   | tgagcctcaagaaacactactctgtc  | 1260 |
| Db | 1201 | AAGTTAAGAGCAATGGAGAAAGACT   | GAGCTCAAGAGCAACTACTCTTGCT   | 1260 |
| OY | 1261 | accatacctctgtctggtgaacccca  | gtaaccccttcgaagccaaagtatgat | 1320 |
| Db | 1261 | ACCATACCTCTGCTGGATGATACCC   | AGTACCCCTTCGGAACCAAAATGAT   | 1320 |
| OY | 1321 | gaacctgagctcgagcccgcaagat   | tgaccttgatctgtagcagagaggaag | 1380 |
| Db | 1321 | GAACTGAGCTCGAGCCCGCAGATG    | AGACTGTGATGAGGACCGAGAGACG   | 1380 |
| OY | 1381 | -----                       | -----                       | 1381 |
| Db | 1381 | ACATTTTGCACCAAGCCCAAGCCT    | CCAGAGCCGTGGGCTGCCACCATCT   | 1440 |
| OY | 1381 | -----                       | -----                       | 1381 |
| Db | 1441 | GTCATGTCACACAGAGCCCAATT     | CTCATCTACAGGCCAGTGTGAAAT    | 1500 |
| OY | 1381 | -----                       | -----gattcc                 | 1387 |
| Db | 1501 | CCACAGAGATGTTGAAGAGAGATG    | GTGTTATCAAGACAGATTCAAGACA   | 1560 |
| OY | 1388 | ggaataatgagaaagctcttgaccct  | ctactctatctctaccccaagagacat | 1447 |
| Db | 1561 | GGAATATGAGGAAGGCTTTGACCC    | CACTATGTTTCAACCCACAGCATCA   | 1620 |
| OY | 1448 | aggaatcccgagctctctacgcata   | aagaagaagagatccttagacctg    | 1507 |
| Db | 1621 | AGGATGTCGGGCTCTACGATCAAA    | GAGAGGAGATCTTTAACCTGGCCCT   | 1680 |
| OY | 1508 | gtctgagctccccaattggaaaggt   | tgctctctctctgtatagaagggagct | 1567 |
| Db | 1681 | GTTGGACTCCCAATGGAGAGTGTG    | CTTCTCTGTGTATAGGGGAGCTGTG   | 1740 |
| OY | 1568 | agcgagctgctgagctatgttbaa    | agggagagctatctgcaatcaacagc  | 1627 |
| Db | 1741 | AGCGGCAATGGTGCAATGTGAAA     | GGGGAGAGATCATGTGCATCAAGG    | 1800 |
| OY | 1628 | cagactacaacctgagctgaggtc    | gagctgagccttcagaagccttgaat  | 1687 |
| Db | 1801 | CAGACTACACCTGGCTGAGGCTG     | AGCGTCCCTCGAGAGGCCCTGGA     | 1860 |
| OY | 1688 | actgagatcgacctgtgtgtgcgct   | ctgtgcccccaagagatatacagat   | 1747 |
| Db | 1861 | ACTGGATCGACTGTGTGTCGTCG     | CTGCCCCCAAGAGATATACATGAC    | 1920 |
| OY | 1748 | tcctgtcgaagtcaccaaaaggga    | aacaaatcaocgcttagaagaaagt   | 1807 |
| Db | 1921 | TCTTGTGGAATCCAAAAGGGAAC     | CAAAATTACCGCGTTAGGAACA      | 1980 |
| OY | 1808 | cccaacctctgtgaacaaagcctc    | tgagcaagccttgaagagccacatga  | 1867 |
| Db | 1981 | CCCAACCTCTGTGAACAAAGCTCT    | GGACAGCCTTGAGAGGCCACATGAC   | 2040 |
| OY | 1868 | gatgtgcatccttgggagccttga    | atctatcacccaaggaattcaaatc   | 1927 |
| Db | 2041 | GATGTGACTCTTTGGAGCTTAAT     | CTATCAACCCAGGATCTCAAACT     | 2100 |
| OY | 1928 | acagagggccaagataaagaaagc    | tcgagcaacttlttbaagccaatgt   | 1987 |
| Db | 2101 | ACCAAGGCGCATATAGGAADAC      | GCTCGGGCACTTTTATAGAGCCAA    | 2160 |
| OY | 1988 | gagcagcagacgcttlttggaaga    | atctcaagaatcagaactctatcct   | 2047 |

|    |      |   |      |
|----|------|---|------|
| Db | 2161 | GAGCAGCAGCCGCTTTGGGAAAGATCTCAGAGATCCAGATCTCTATTTCTTCTCTCGG    | 2220 |
| Qy | 2048 | cccaatgaatttggctctctcccaagctttggggactctcttccttgaacctataaagacc | 2107 |
| Db | 2221 | CCCATGATTTGGTCTCTCCACAGCTTTGGGGACCTTCTTGACCCCTATTAAGAC        | 2280 |
| Qy | 2108 | ccactggagctctctctctccacaccctctccctctgcctctgcctctaaatctgcagag  | 2167 |
| Db | 2281 | CCACGGGAGTCTCTCTCTCTCCATCCCTCTCCCTCTGCTGCTATTTCTCTCCAGG       | 2340 |
| Qy | 2168 | attgtcaactccaacacttactctgagctcatctataataaataacagattatttccagc  | 2227 |
| Db | 2341 | ATTGTCATCTCAAACTTACTCTAGCTCTTATATATATAAACAAGATTATTATTCAGC     | 2400 |
| Qy | 2228 | ttaaaaaa 2236   |      |
| Db | 2401 | TTAAAAAA 2409   |      |

RESULT 6  
 X40060 standard; DNA, 1306 BP.  
 AC X40060;  
 DT 02-JUL-1999 (first entry)  
 DE Colon cancer associated gene.  
 KW Cancer associated antigen; diagnosis; research; treatment; human;  
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;  
 KW prostate cancer; ss.  
 OS Homo sapiens.  
 PN W09J04265-A2.  
 PD 28-JAN-1999.  
 PF 15-JUL-1998; U14679.  
 PI 22-JUN-1998; US-102322.  
 PR 17-JUL-1997; US-896164.  
 PR 10-OCT-1997; US-061599.  
 PR 10-OCT-1997; US-061765.  
 PR 10-OCT-1997; US-948705.  
 PR 11-OCT-1997; GB-021697.  
 PA (LUDMW-) LUDMWG INST CANCER RES.  
 PI Chen Y, Gout I, Gure A, Ohare M, Odate Y, Old LJ,  
 PI Pfeundschnu M, Sahn U, Scanlan MC, Stockert E,  
 PI Tureci O;  
 DR WPI; 99-132448/11.  
 PT New isolated cancer associated nucleic acids and polypeptides -  
 PT isolated using sera from cancer patients, used to develop products  
 PT for the diagnosis, monitoring or treatment of cancers  
 PS Claim 67; Page 662; 787pp; English.  
 CC The invention relates to a method for diagnosing a disorder characterised  
 CC by expression of a human cancer associated antigen precursor coded for by  
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a  
 CC biological sample isolated from a subject with an agent that specifically  
 CC binds to the NAM, an expression product or a fragment of an expression  
 CC product complexed with an HLA molecule; and (b) determining the  
 CC interaction between the agent and the NAM or the expression product as a  
 CC determination of the disorder. The products and methods can be used in  
 CC the diagnosis, monitoring, research, or treatment of conditions  
 CC characterised by the expression of various cancer associated antigens.  
 CC The invention provides nucleic acid sequences and encoded polypeptides  
 CC which are cancer associated antigen precursors expressed in human breast  
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and  
 CC lung cancer.  
 CC Sequence 1306 BP; 348 A; 331 C; 375 G; 252 T;

| Query Match           | 58.4%          | Score 1306  | DB 1     | Length 1306 |
|-----------------------|----------------|---|----------|-------------|
| Best Local Similarity | 100.0%         | Pred. No. 0   |          |             |
| Matches 1306          | Conservative 0 | Mismatches 0  | Indels 0 | Gaps 0      |
| Oy                    | 931            | aaatatagccgagcgtctaccatctctcattttagctgacatctgcccgaagctgttcacatg | 990      |             |
| Db                    | 1              | AAAATACCCGCGAGCCTGACATCTTCATTGATGTCGACGCTGCGCCGGAGACTGTTCATG    | 60       |             |

|    |      |   |      |
|----|------|---|------|
| OY | 991  | acgagccgggagcgcgctcgcgagagcgcgcgacgctgagctgcagcgcgagagctctc     | 1050 |
| Db | 61   | ACAGACCGGGAGCGGCTGGGAGAGGCGCGGACCGTGAGCTCGAGCGGAGAGCTTCTC       | 120  |
| OY | 1051 | atgcagaagcgcgctcgcgatctgagatcccaaaagatctccaggaagcaagagagatgag   | 1110 |
| Db | 121  | ATGCAGAAAGCGGCTGGCGATGGATCCACACAGATCTCTCAGAGCAGCAGATGTGAG       | 180  |
| OY | 1111 | cggcgaaggaagaaagaatctgcgcaagagcgagcagagaaatctgagatctacggaag     | 1170 |
| Db | 181  | CGCGAAAGGAGAAAGAAATTGGCCAGAAAGCAGCAGAGAGAAATGAGATACCGGAG        | 240  |
| OY | 1171 | gagatggaacagatctgtagagaggaagagaagttcaagaagcaatggyaagaagctg      | 1230 |
| Db | 241  | GAGATGGAAACGATTGTAGAGGAGGAAGAACATTTAACACCAATGGGAACAACTGG        | 300  |
| OY | 1231 | ggctcaaggaagcagctaacctctgcccataaacaatacgcctgagatgaacccagctacc   | 1290 |
| Db | 301  | GGCTCAAAAGACACGCTACTCTTGCTTAACCATACTGCTGAGGTACACCCATTACC        | 360  |
| OY | 1291 | cttcgcaagccaaagatctgaacagagctggaacccgagctcgagcccgagatgacctg     | 1350 |
| Db | 361  | CTTGCAAGCCAAAGTGTATGATCAGGGAGTGGAACGTGACTGAGCGCCGACATGACCTG     | 420  |
| OY | 1351 | gatgagagcacggagagcgagagcagagatctccggaatatgaggaagcttgac          | 1410 |
| Db | 421  | GATGAGAGCACGGAGGAGCAGGGAGAGACAGATTTCCGAAATATGAGGAAGGCTTTGAC     | 480  |
| OY | 1411 | ccctacactatgttcaccccaagacagatcatggygaagatgctcogtctctaagctc      | 1470 |
| Db | 481  | CCCTACTCTATATGTTACCCACGAGACAGATCATGGGAAGATGTCCGGCTCTTACGATWC    | 540  |
| OY | 1471 | aagaagagagatctcttaagaccctggcccttgaaagcggtgtgtaactccccatitggaag  | 1530 |
| Db | 541  | AAGAAGAGAGGATCTTATGACTGTGCCCTTGGAAGCGGCTGTGACTCCCCCTTTGGAGAG    | 600  |
| OY | 1531 | gtgtgcgttctcgtctgtgtaagcggygagctgctgtagcgycatgtgtgcatgtgtaa     | 1590 |
| Db | 601  | GTGTCGTTTCTGCTGTATGAGACGGGAGCTGTGACGCGCATGTGNGCATTTGTGAAA       | 660  |
| OY | 1591 | ggggaagagatcatggygaatccaacgagagatgtgtaacagactaaccttgctgtaggct   | 1650 |
| Db | 661  | GGGGACGAGATCATGGCAATCAACGGCAAAATGTACAGACTTACACCTGTGCTAGGCT      | 720  |
| OY | 1651 | gaagctgcccgcgaagagagcctggaatcaagggcgagagatgtagatcgaactgtgtgttc  | 1710 |
| Db | 721  | GACGCTGCCCTGCGAAGAGCCCTGGAATCAGGGCGGGGACTGGAATGCACTTGTGTTGCC    | 780  |
| OY | 1711 | gtctgcccccaaaagagatctgacgatgagcttgacacttctgtctaagtccaagaaggya   | 1770 |
| Db | 781  | GTCGTCCCCCAAGGAGTATGACGATGAGACTGACTTGTGCTGATGAATCCAAAAGGGGA     | 840  |
| OY | 1771 | aaccaaattcacgqgttaagaaacagttagctccggcccccaactcgttgacaacaagcct   | 1830 |
| Db | 841  | AACCAAAATTCACGGGTTAGGAAACAGTAGACTCTCGGCCACCTCGTGAAACAAAGCCT     | 900  |
| OY | 1831 | cgagcagagccttgagagagcgacaatgacacacacagaatgcatcactcttggaagactgat | 1890 |
| Db | 901  | CGGACACGACTTTGAGAAAGGCCACATGACACACACAGATGGATCTTGGAACTGAT        | 960  |
| OY | 1891 | ctataccacagaaatctccaactccctcttgcccttgaaacacagggccagaataagagaagc | 1950 |
| Db | 961  | CTATACCCACAGATCTCAAAATCCCTTGGCCCTAACAAGGCCAAGATAAGAAACAGC       | 1020 |
| OY | 1951 | tcggggcaactttttgaagccaatgtgaggaagagagcgacgacgctttggyagaa        | 2010 |
| Db | 1021 | TCGGGCGCACTTTTGAAGGCCAATGTGGAGAAAGGAGCAGCCAGCCGTTTGGAGAA        | 1080 |
| OY | 2011 | gactcagaagatccaagactcaactctctccctctgcccagaatgaaattgtgctctcca    | 2070 |
| Db | 1081 | GATCTCAAGATCCAGACTCTCATCTCTTCCCTCTGACCAGTGAATTTGGCTCTCCCA       | 1140 |
| OY | 2071 | gctttggygagctctctccttgaaaccttaataagaccacatgagatctctcctctca      | 2130 |

|   |       |  |       |
|---|-------|--|-------|
| Dd  | 1141  | GGTTTGGGAGACTCTTCCTTGACCCATAATGAACCCACGCGAGTCTCTCTCACA           | 1200  |
| Oy  | 2131  | tccctcctcgcgccctcgctctaattgctgcgaagatgtcaactccaacttaact          | 2190  |
| Dd  | 1201  | TCCCTCTCTCTGCCTCTGTCTTAATTCTCTGCAAGATTGTCACTCCAACTTACTCT         | 1260  |
| Oy  | 2191  | gaagtcataataaataaacagatttatttccacgttaaaaaaa                      | 2236  |
| Dd  | 1261  | GAGCTCATTAATAAATAAACAGATTATTTCAGCTTAATAAAAAA                     | 1306  |
| <br>RESULT 7<br>v73805/c<br>ID v73805 standard; DNA: 32207 BP.<br>AC V73805;<br>DT 25-FEB-1999 (first entry)<br>DE KSHV LUR DNA (nucleotides 105,301-137,507).<br>KW Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;<br>KM dIlydrofolate reductase; LOR; long unique region; vaccine; prophylaxis;<br>KN diagnosis; treatment; HHV8; capsid protein IV; tegument protein IV;<br>KV glycoprotein; kaposin; cyclin D; immediate early protein; IEp; OX-2;<br>KM v-abl; G-protein coupled receptor; FGARAT; ds.<br>OS Kaposi's sarcoma-associated herpesvirus.<br>PN US5849564-A.<br>PF 15-DEC-1998:<br>PR 29-NOV-1996: 770379.<br>PA (UYCO) UNIV COLUMBIA NEW YORK.<br>PI Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;<br>PT Kaposi's sarcoma-associated herpes virus nucleic acid - encodes<br>PR dIlydrofolate reductase and is useful for treatment, prophylaxis<br>PT or diagnosis of Kaposi's sarcoma<br>PS Disclosure: Column 155-182; 109pp: English.<br>P5 This sequence is a fragment of the Kaposi's sarcoma-associated<br>CC herpesvirus (KSHV) LUR (long unique region). This fragment contains<br>CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67<br>CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,<br>CC ORF69, K12 which encodes kaposin, K13, ORF72 which encodes cyclin D,<br>CC ORF73 which encodes immediate early protein (IEP), K14 which encodes<br>CC OX-2 (v-abl), ORF74 which encodes G-protein coupled receptor, ORF75<br>CC which encodes tegument protein/FGARAT, K15. KSHV is a new human<br>CC Herpesvirus (HHV) believed to cause Kaposi's sarcoma (KS) which is the<br>CC most common form of neoplasm occurring in persons with acquired immune<br>CC deficiency syndrome (AIDS). The DHFR protein is useful for vaccination,<br>CC prophylaxis, diagnosis and treatment of a subject with Kaposi's sarcoma<br>CC and for detecting expression of a DNA virus associated with Kaposi's<br>CC sarcoma in a cell.<br>SQ Sequence 32207 BP; 7229 A; 9156 C; 8713 G; 7109 T; |       |  |       |
| <br>Query Match 3.1%; Score 68.4; DB 1; Length 32207;<br>Best Local Similarity 47.6%; Pred. No. 9,6e-08;<br>Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;  |       |  |       |
| Oy  | 979   | gagctgttatgaaccaggaccgggagcggtcggcaagggcgcgcaagcgtagctgcaagcg    | 1038  |
| Dd  | 19815 | GAGCACACACACACAGATGATGACGACGACGACAGATGAGACACACAGATGACGAG         | 19756 |
| Oy  | 1039  | caggagcttcatactgacaagaacggcgctcgggatggagtccaaacaagatctccagagcag  | 1098  |
| Dd  | 19755 | CAGCAGCAGAGATGACACGACGACGACGACGACGACGACGACGACGACGACGACGACG       | 19696 |
| Oy  | 1099  | caggagatggagcgggcaaaagaaaagaattgcccaagaagcgcaagaggaanaatggag     | 1158  |
| Dd  | 19695 | GAGTTTAGAGACACAGACGACGACGAGTTAGAGATCAGAGACGACGAGTTAGAGACGAG      | 19636 |
| Oy  | 1159  | agatccggaagaagatatggaacagatttggaaggagggaagaagaatttaagaagcaatgg   | 1218  |
| Dd  | 19635 | CAGGAGTTTAGAGACGACGACGACGAGCTTAGAGAGACGACGACGAGTTTAGAGAGACGAG    | 19576 |
| Oy  | 1219  | gaagaagaactctggggctcaagaagaaacgctacctctgtccctaacaacatcaactctagga | 1278  |

|      |      |   |   |       |
|------|------|---|---|-------|
| D    | b    | 19575   | GAGCAGGAGTTTAGACGACGAGCAGCGAGCTTATAGAGCAACGACGACGAGCTTATAGAGCAG | 19516 |
| Oy   | 1279 | caccacgtaccccttcgcgaagcacaagtatgatctaggggagltgaaacctgagctcgagccc        | 1338  |       |
| D    | b    | 19515   | CAGGACGACGAGATTAGAGGACGACGACGAGGATTAGAGGACGACGAGGAGTGGAAGACCA   | 19456 |
| Oy   | 1339 | gcagatgacctgatatgtaggcaagcagaagcagaagagacgaattccggaataatag              | 1398  |       |
| D    | b    | 19455   | GAGCAGGAGGTGGAAAGACCAAGCAGCAGGACGAGCAAGAAGATTTAGAGAGGTGGAG      | 19396 |
| Oy   | 1399 | ga 1400   |   |       |
| D    | b    | 19395   | GA 19394  |       |
| <br> |      |   |   |       |
| R    | E    | S   |   |       |
| I    | D    | V19941/C  |   |       |
| A    | C    | V19941;   |   |       |
| D    | E    | 03-AUG-1998 (first entry)   |   |       |
| D    | R    | KSHV long unique coding region and terminal repeat.                     |   |       |
| K    | M    | KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II:   |   |       |
| K    | M    | interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis |   |       |
| K    | M    | complement-binding protein; glycoprotein; capsid protein IV; infection; |   |       |
| K    | M    | lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides |   |       |
| K    | M    | HIV immune status; anti-inflammatory agent; therapy; ds.                |   |       |
| K    | M    | Kaposi's sarcoma-associated herpes virus.                               |   |       |
| F    | H    | Key location/Qualifiers   |   |       |
| F    | T    | CDS   | 1142..2794  |       |
| F    | T    | /tag= a   |   |       |
| F    | T    | /product= complement-binding protein                                    |   |       |
| F    | T    | 8699..11236   |   |       |
| F    | T    | /tag= b   |   |       |
| F    | T    | /product= glycoprotein B  |   |       |
| F    | T    | complement (17261..17875)   |   |       |
| F    | T    | /tag= c   |   |       |
| F    | T    | /product= interleukin 6   |   |       |
| F    | T    | complement (21548..21832)   |   |       |
| F    | T    | /tag= d   |   |       |
| F    | T    | /product= macrophage inflammatory protein II                            |   |       |
| F    | T    | complement (27137..27424)   |   |       |
| F    | T    | /tag= e   |   |       |
| F    | T    | /product= interferon regulatory factor 1                                |   |       |
| F    | T    | 28661..29741  |   |       |
| F    | T    | /tag= f   |   |       |
| F    | T    | /product= protein T1.1  |   |       |
| F    | T    | complement (58976..60175)   |   |       |
| F    | T    | /tag= g   |   |       |
| F    | T    | /product= glycoprotein M  |   |       |
| F    | T    | complement (69412..69915)   |   |       |
| F    | T    | /tag= h   |   |       |
| F    | T    | /product= glycoprotein L  |   |       |
| F    | T    | complement (88410..88910)   |   |       |
| F    | T    | /tag= i   |   |       |
| F    | T    | /product= interferon regulatory factor 2                                |   |       |
| F    | T    | 89600..90541  |   |       |
| F    | T    | /tag= j   |   |       |
| F    | T    | /product= interferon regulatory factor 3                                |   |       |
| F    | T    | 90173..90643  |   |       |
| F    | T    | /tag= k   |   |       |
| F    | T    | /product= glycoprotein X  |   |       |
| F    | T    | complement (93636..94127)   |   |       |
| F    | T    | /tag= l   |   |       |
| F    | T    | /product= interferon regulatory factor 4                                |   |       |
| F    | T    | complement (111931..112443)   |   |       |
| F    | T    | /tag= m   |   |       |
| F    | T    | /product= capsid protein IV   |   |       |
| F    | T    | complement (123808..127296)   |   |       |
| F    | T    | /tag= n   |   |       |
| F    | T    | /product= immediate early protein                                       |   |       |
| N    |      | WO9804576-A1.   |   |       |

[illegible]



CC nucleic acid (see also V21568-86) into a cell. The polycationic  
CC agent can condense with the nucleic acid and inhibit serum and/or  
CC nuclease degradation of the nucleic acid. The nucleic acid can be  
CC a vector, may express a therapeutic protein or a vaccinating viral  
CC or cancer antigen, or is itself therapeutic (antitense or  
CC ribozyme). The methods and compositions can be used in the gene  
CC therapy of many diseases.  
SQ Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T;

|                           |       |                   |          |             |
|---------------------------|-------|-------------------|----------|-------------|
| Query Match               | 2.6%  | Score 58          | DB 1     | Length 9600 |
| Best Local Similarity     | 46.9% | Pred. No. 2.8e-05 |          |             |
| Matches 181, Conservative | 0     | Mismatches 205    | Indels 0 | Gaps 0      |

|    |      |  |      |
|----|------|--|------|
| QY | 997  | cggagacgagcttgcgcagagcgcgcttgagcttcagcgacaggaattcttcattgaag        | 1056 |
|    |      |  |      |
| Db | 1016 | CAGGAGCAGAGGAGGAGGGCCAGGAGCCAGAGAGGGCCAGAGAGGGCCAGGACGAC           | 1075 |
| QY | 1057 | aagcgagcttgcgcgttgagttcccaacaagatctctccagagcaagcaggaatattgagcgccaa | 1116 |
|    |      |  |      |
| Db | 1076 | GAGGGCCAGGAGCAGAGGAGAGGGGCCAGAGAGGGGCAAGAGGGCCAGGACAGGAGGGCCAG     | 1135 |
| QY | 1117 | aggaagaaagaaatttcccagaaagcagcagagagaaatltagagattaccggaagagatg      | 1176 |
|    |      |  |      |
| Db | 1136 | GAGCAGAGCGGCGAGGAGGCGAGGAGGGGCGAGGAGCAGAGAGGGCCAGAGGCGAGG          | 1195 |
| QY | 1177 | gaacagattttagagagagagaaagaggtttaagaagcaattggaagaagacttggggtcca     | 1236 |
|    |      |  |      |
| Db | 1196 | GAGGGCGAGGAGGGGCGAGGCGAGGAGGGGCGAGAGGGGCGAGAGAGAGGGGCGAG           | 1255 |
| QY | 1237 | aaggaacagctactcttgccttaaacatcatctgtctgaagtatcacccagtaacctcttgc     | 1296 |
|    |      |  |      |
| Db | 1256 | GAGGGCGAGGAGGAGGAGGAGGGGCGAGAGGGGCGAGAGAGAGGGGCGAGAGGGGCGAG        | 1315 |
| QY | 1297 | aagccaaagtatgatcagggagttgaaaccttgaagcttcgagcccgacagatgaccttgatga   | 1356 |
|    |      |  |      |
| Db | 1316 | GAGCAGAGAGGGCGAGGAGGGGCGAGGAGCGAGAGGGGCGAGAGGGGCGAGAGGAGAG         | 1375 |
| QY | 1357 | ggcacggagagagcagggagagacggga                                       | 1382 |
|    |      |  |      |
| Db | 1376 | GGGCAGAGCAGAGGAGGGGCGAGGAGCA                                       | 1401 |

[illegible]

```

FT      /tag= g
FT      /label= EBNA-1
FT      4190. .6374
FT      /tag= h
FT      /function= orip
FT      4295. .4887
FT      /tag= i
FT      /note= "family of repeats"
FT      5866. .5978
FT      /tag= j
FT      /note= "dyad region"
FT      6375. .6457
FT      /tag= k
FT      /label= HSV-TK_terminator_3'-end
FT      6975. .7975
FT      /tag= l
FT      /phenotype= neomycin_resistance
FT      /note= "rns neomycin phosphotransferase gene"
FT      7975. .8112
FT      /tag= m
FT      /label= TK_promoter
FT      8114. .8594
FT      /tag= n
FT      /function= M13_orf1
FT      8595. .10414
FT      /tag= o
FT      /label= delta_2a
FT
FT      WO9323541-A.
FT      25-NOV-1993.
FT      17-MAY-1993; US-04648.
FT      18-MAY-1993; US-884811.
FT      18-MAY-1992; US-885971.
FT      (GETH ) GENENTECH INC.
FT      PA
FT      Godowski PJ, Locker NA, Mark MR;
FT      WPI; 93-386573/48.
FT      DR
FT      Hepatocyte growth factor variants - are resistant to proteolytic
FT      cleavage into its two-chain form, used to treat malignancies
FT      associated with HGF receptor
FT      Example 1; Fig 6; 87pp; English.
FT      Plasmid pcisEBON (a PKR5 derivative) is an episomal CMV driven
FT      expression plasmid. HuHGF variants with enhanced receptor binding
FT      activity were produced by site-directed mutagenesis. Stable
FT      populations of preferred HGF variants were obtained by transfecting
FT      human embryonic kidney 293 cells and then these were subcloned in
FT      pcisEBON. See R52940-R52949 for examples of pref. HGF variants.
FT      Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T.
FT

```

[illegible]

| Query Match | Best Local Similarity | Matches | Score | DB | Length |
|-------------|-----------------------|---------|-------|----|--------|
| 997         | 46.9%                 | 181     | 58    | 1  | 10596  |
| 2550        | 0                     | 0       | 205   | 0  | 0      |
| 1057        | 0                     | 0       | 205   | 0  | 0      |
| 2610        | 0                     | 0       | 205   | 0  | 0      |
| 1117        | 0                     | 0       | 205   | 0  | 0      |
| 2670        | 0                     | 0       | 205   | 0  | 0      |
| 1177        | 0                     | 0       | 205   | 0  | 0      |

| QY   | 1237  | aaggaacagctactcttctgccttaaaacatcactctgtagtgaaccacccagttacccttcgc | 1296 |
|--|---|--|------|
| DB   | 2730  | GAGGGCGCAGAGGGCGCAGGACGAGGGGCGCAGAGGGGCGCAGGACGAGGAGGAGGGCGAG    | 2789 |
| QY   | 1237  |  |      |
| DB   | 2790  | GAGGGCGCAGGACGACGAGGAGGGCGCAGGAGGGCGCAGGACGAGGAGGAGGGCGAG        | 2849 |
| QY   | 1297  | aagccaagtatgatcaggaggttgaaacttgatcctcgagccgcagatgacttgatbga      | 1356 |
| DB   | 2850  | GAGCGAGGAGGGCGCAGAGGGCGCAGGAGGAGGGCGCAGGAGGGCGCAGGAGGAGGAG       | 2909 |
| QY   | 1357  | gccaagagagagcaggagagacagga                                       | 1382 |
| DB   | 2910  | GGCGAGGAGCAGAGGAGGGCGCAGGAGCA                                    | 2935 |
| RESULT 13  |   |  |      |
| ID   | X15650  | standard; DNA; 10596 BP.   |      |
| AC   | X15650:   |  |      |
| DT   | 10-MAY-1999   | (first entry)  |      |
| DE   | Nucleotide sequence of plasmid PCIS_EBON.                                 |  |      |
| KW   | Plasmid PCIS_EBON; hepatocyte growth factor; HGF; variant;                |  |      |
| OS   | HGF receptor activation; sa.  |  |      |
| KS   | Synthetic.  |  |      |
| PN   | US5879910-A.  |  |      |
| PD   | 09-MAR-1999.  |  |      |
| PF   | 09-FEB-1994; 194087.  |  |      |
| PR   | 09-FEB-1994; US-194087.   |  |      |
| PA   | (GEMH ) GENENTECH INC.  |  |      |
| PI   | Godowski PJ, Lokerer NA, Mark MR;   |  |      |
| DR   | WPI; 99-203949/17.  |  |      |
| PT   | New hepatocyte growth factor variants - useful for studying               |  |      |
| PT   | structure-function relationships in the wild type molecule and for        |  |      |
| PT   | treating conditions associated with chronic hepatocyte growth factor      |  |      |
| PT   | receptor activation   |  |      |
| PS   | Example 1: Fig 6A-F; 40pp; English.                                       |  |      |
| CC   | The present sequence represents the nucleotide sequence of PCIS_EBON,     |  |      |
| CC   | which is used in the course of the invention. The specification           |  |      |
| CC   | describes a hepatocyte growth factor (HGF) variant (HGfV) comprising an   |  |      |
| CC   | amino acid (aa) alteration at or adjacent to position 692 of the          |  |      |
| CC   | wild-type human HGF (huHGF) aa sequence. HGfV may be used in studies to   |  |      |
| CC   | identify the structure-activity relationships of HGF in order to identify |  |      |
| CC   | the functionally important domains in the aa sequence. It may also be     |  |      |
| CC   | used to identify aa residues which are responsible for the interaction of |  |      |
| CC   | HGF with its receptor, and those aa that are responsible for the          |  |      |
| CC   | biological activity of it. Variants of HGF which have enhanced receptor   |  |      |
| CC   | binding affinity (compared to wild-type huHGF) and are more biologically  |  |      |
| CC   | active than wild type huHGF, may be used as huHGF agonists. Conversely,   |  |      |
| CC   | variants of HGF which have enhanced receptor binding affinity (compared   |  |      |
| CC   | to wild-type huHGF) but which are biologically inactive may be used as    |  |      |
| CC   | huHGF antagonists, and may be used to block the binding of wild-type      |  |      |
| CC   | huHGF to its receptor. This permits the treatment of pathogenic           |  |      |
| CC   | conditions associated with the activation of an HGF receptor, such as     |  |      |
| CC   | malignancies associated with chronic HGF receptor activation.             |  |      |
| CC   | HGfV-immunoglobulin chimeras may be produced (by standard methods) and    |  |      |
| CC   | used in protein A purification, immunohistochemistry, and                 |  |      |
| CC   | immunoprecipitation techniques (in place of anti-HGF antibodies) or in    |  |      |
| CC   | screening studies to identify inhibitors of HGF-HGF interactions.         |  |      |
| SO   | Sequence  | 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;                        |      |
| Query Match: 2.6%; Score 58; DB 1; Length 10596;               |   |  |      |
| Best Local Similarity 46.9%; Pred. No. 3e-05;                  |   |  |      |
| Matches 181; Conservative 0; Mismatches 205; Indels 0; Gaps 0; |   |  |      |
| QY   | 997   | cggagcagcgcttgcagcagagcgcgacgctgcagctgcagcggcagagcttcattgcag     | 1056 |
| DB   | 2550  | CAGGAGCAGAGAGGAGGCGCAGGAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGCAG       | 2609 |
| QY   | 1057  | aagcgctcggcgtagtgcacaagaatctccacggagcagcagagagatgtagcgcgca       | 1116 |
| DB   | 2610  | GAGGGGCGAGCAGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAGGAGGCGCAG    | 2669 |



| Accession | Sequence   | Position |
|-----------|--|----------|
| QY 1117   | agsgaaagaatctgcccgaagcgacgcaagaataatgtgagatctccgaaagagatg      | 1176     |
| Db 2670   | GAGCAGAGGGGCGAGACAGAGAGGGGCGAGACAGAGGGCGAGAGGGCGAGAGCAG        | 27299    |
| QY 1177   | gaacagatctgtatgagagagagaagaagtttaagaagcaatgtggaagaacttgygtcta  | 1236     |
| Db 2730   | GAGGGCGAGAGGGCGCAGAGCGAGAGGGGCGAGAGAGGCGACAGAGAGAGGGCGAG       | 27899    |
| QY 1237   | aaggaacagctactcttgccttaaacatcatctgcgttgagtgatacaaccagtacccttgc | 1286     |
| Db 2790   | GAGGGGCGAGACAGAGAGAGAGGGGCGAGAGGGGCGAGACAGAGAGGGCGAGGGCGAG     | 2849     |
| QY 1297   | aagcacaagatattgatcaggtgaggttggaaacctgagctcgcgcagattgaccttgatga | 1356     |
| Db 2850   | GAGCAGAGAGGGGCGAGAGGGGCGAGAGCGAGAGGGGCGAGAGGGGCGAGAGGAGAG      | 2909     |
| QY 1357   | ggcagcgagagcagcgagagcgacga                                     | 1382     |
| Db 2910   | GGGCAGAGCAGAGGGGCGAGAGGCA                                      | 2935     |

|        |   |
|--------|---|
| RESULT | 14  |
| ID     | V55830  |
| AC     | V55830 standard; DNA; 795 BP.   |
| DT     | 18-NOV-1998 (first entry)   |
| DE     | FIGA insert stabilising polypeptide encoding DNA.                         |
| KW     | Fusion protein; stabilising polypeptide; proteolytic degradation;         |
| KW     | resistance; half-life; autoimmune disease; inflammation; nitro drug;      |
| KW     | IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;   |
| KW     | nitroreductase protein; enzyme therapy; prodrug therapy; protease;        |
| KW     | cancer; pathological condition; ss.                                       |
| OS     | Epsilon-barri virus.  |
| FT     | Key   |
| FT     | CDS   |
| FT     | Location/Qualifiers   |
| FT     | 1..788  |
| FT     | /tag= a   |
| FT     | /product= "stabilising polypeptide"                                       |
| PN     | W09822577-A1.   |
| PD     | 28-MAY-1998.  |
| PF     | 17-NOV-1997; IB1508.  |
| PR     | 25-JUN-1997; US-048945.   |
| PR     | 15-NOV-1996; US-030986.   |
| PA     | (MASU/) MASUCCT M G.  |
| PI     | Masucci MG;   |
| DR     | WPI; 98-312463/27.  |
| DR     | p-PSDb; W79128.   |
| PT     | New fusion proteins resistant to proteolytic degradation -                |
| PT     | comprising a core protein with a stabilising polypeptide comprising       |
| PS     | a peptide sequence containing glycine repeats                             |
| PS     | [Disclosure; Fig 3; 120pp] English.                                       |
| CC     | This DNA encodes a stabilising polypeptide and is the FIGA insert of the  |
| CC     | Invention. The invention provides a method for increasing the resistance  |
| CC     | of a core protein to proteolytic degradation that comprises linking or    |
| CC     | inserting onto or into the core protein a stabilising polypeptide of      |
| CC     | formula [(Gly)X(Glyb)Y(glyc)Z] <sub>n</sub> where Gly, Glyb, Glyc are 1-6 |
| CC     | sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,     |
| CC     | Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not   |
| CC     | be identical from n repeat to n repeat. Alternatively a nucleic acid      |
| CC     | encoding the stabilising polypeptide can be linked onto or inserted into  |
| CC     | a nucleic acid encoding a core protein. The fusion proteins of the        |
| CC     | invention are more resistant to degradation by proteases and, thus, have  |
| CC     | a longer half-life than the unfused core protein. The products can be     |
| CC     | used for treating autoimmune diseases, cancer and inflammation. In        |
| CC     | particular, the core protein may be an IkappaB regulator protein for the  |
| CC     | treatment of inflammatory bowel disease, or a nitroreductase protein      |
| CC     | which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  |
| CC     | or other pathological conditions. The fusion proteins can also be used in |
| CC     | diagnostic methods such as in vivo imaging.                               |
| QJ     | Sequence 795 BP; 200 A; 104 C; 478 G; 13 T;                               |

|                       |        |           |          |             |
|-----------------------|--------|-----------|----------|-------------|
| Query Match           | 2.58;  | Score 57; | DB 1;    | Length 795; |
| Best Local Similarity | 51.8%; | Pred. No. | 1.5e-05; |             |

[illegible]

RESULT 15  
ID Q23092  
AC Q23092; standard: DNA; 543 BP.  
DT 17-AUG-1992 (first entry)  
DE Antigen tc-7a gene.  
KW Oocysts; Mab 12-07; sporozoite; ss.  
OS Elmeria tenella.  
PN WO9204460-A.  
PD 15-MAR-1992.  
PF 03-SEP-1991; U06430.  
PR 12-SEP-1990; US-581693.  
PA (GENE-) GENEX Corp.  
PI Jacobson JW, Strausberg RL, Wilson SD, Pope SH, Strausberg SL;  
PI Ruff MD, Augustine PC, Danforth HD;  
PI WPI: 92-114365/14.  
DR P-PSDB; R22392.  
PT Vaccinate against avian coccidiosis - comprising recombinant  
PT Elmeria antigen ac-1b or ac-6b gene, or microorganisms expressing  
PT them  
PS Claim 5; Page 39 + Fig 5; 56pp; English.  
CC To identify antigens of *E. tenella*, expression libraries were  
CC prep'd. in lambda vector, lambda gfil, using cDNA prep'd. from polyA  
CC mRNA isolated from *E. tenella* oocysts. The cDNA expression  
CC library was screened with monoclonal antibody (Mab) 12-07 which was  
CC raised against the sporozoite stage of *E. tenella*. The library  
CC to be screened was plated on a host that allows lysis and plaque  
CC formation. During induction of the antigens encoded by the phage,  
CC the plaques were identified by screening the filters with Mab 12-07  
CC The cDNA inserts from the Mab 12-07 positive phage were cloned into  
CC bacteriophage M13 and subjected to sequence analysis. Following  
CC sequence analysis, *E. tenella* antigen tc-7a was identified.  
SQ Sequence 543 BP; 228 A; 135 C; 161 G; 18 T.

[illegible]



Mon May 22 16:59:56 2000

us-09-502-945-4.rng

Page 15

```

OY  1175  tgcacagattgttagagaggaagaagtttaagaagca 1214
      |   |   |   |   |   |   |   |   |   |
DB  321  AACGAACCGAAGGAGCAGACGAGCAGCAGCAGAGCA 360

```

Search completed: May 20, 2000, 00:41:50  
Job time: 12162 sec

**THIS PAGE BLANK (user)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:50:55 ; Search time 6491.04 Seconds

(without alignments)  
-335.102 Million cell updates/sec

Title: US-09-502-945-4

Perfect score: 2236

Sequence: 1 cctggcccgctgcgctgc.....tatttccagcttaaaaaa 2236

Scoring table: IDENTITY\_NUC

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_ov:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_com:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_hlg1:\*  
33: gb\_hlg2:\*  
34: gb\_hlg3:\*  
35: gb\_hlg4:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_hlg3:\*  
42: gb\_hlg4:\*  
43: gb\_hlg5:\*  
44: gb\_hlg6:\*

45: gb\_hlg7:\*  
46: em\_hlg1:\*  
47: em\_hlg2:\*  
48: em\_hlg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_hlg8:\*  
52: gb\_hlg9:\*  
53: gb\_hlg10:\*  
54: gb\_hlg11:\*  
55: gb\_hlg12:\*  
56: gb\_hlg13:\*  
57: gb\_hlg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score  | Query Match | Length | DB ID         | Description         |
|------------|--------|-------------|--------|---------------|---------------------|
| 1          | 2236   | 100.0       | 2236   | 40 AF039700   | AF039700 Homo sapi  |
| 2          | 2139.4 | 95.7        | 2213   | 9 AB018687    | AB018687 Homo sapi  |
| 3          | 2078   | 92.9        | 2162   | 40 AF039699   | AF039699 Homo sapi  |
| 4          | 2001.6 | 89.5        | 2283   | 9 AB006955    | AB006955 Homo sapi  |
| 5          | 437.4  | 19.6        | 161190 | 11 AC005137   | AC005137 Human Chr  |
| 6          | 105    | 4.7         | 463    | 13 G55504     | G55504 SHGC-100852  |
| 7          | 81.4   | 3.6         | 7218   | 5 I66494      | I66494 Sequence 14  |
| 8          | 69.6   | 3.1         | 180385 | 40 AC007461   | AC007461 Homo sapi  |
| 9          | 69.2   | 3.1         | 151428 | 55 AC016063   | AC016063 Homo sapi  |
| 10         | 68.8   | 3.1         | 166171 | 42 AC011767   | AC011767 Homo sapi  |
| 11         | 68.4   | 3.1         | 3489   | 16 KSU52064   | U52064 Kaposi's sa  |
| 12         | 68.4   | 3.1         | 32207  | 5 AR065852    | AR065852 Sequence   |
| 13         | 68.4   | 3.1         | 137508 | 16 KSU75698   | U75698 Kaposi's sa  |
| 14         | 67.8   | 3.0         | 133661 | 16 U93872     | U93872 Kaposi's sa  |
| 15         | 67     | 3.0         | 28559  | 16 AF148805   | AF148805 Kaposi's   |
| 16         | 66.4   | 3.0         | 37918  | 41 AC011553   | AC011553 Homo sapi  |
| 17         | 66.2   | 3.0         | 13095  | 9 HSFGR4G     | Y13901 Homo sapien  |
| 18         | 65.8   | 2.9         | 451    | 12 MUSIR3E2   | M10668 Mouse DNA    |
| 19         | 65.2   | 2.9         | 2902   | 11 HSM00885   | AL110228 Homo sapi  |
| 20         | 65     | 2.9         | 97348  | 12 AF091216   | AF091216 Mus muscu  |
| 21         | 65     | 2.9         | 149886 | 12 MMHC438M12 | AF049850 Mus muscu  |
| 22         | 64.8   | 2.9         | 16945  | 44 AC019774   | AC019774 Drosophi   |
| 23         | 63.4   | 2.8         | 195385 | 43 AC016818   | AC016818 Homo sapi  |
| 24         | 63.2   | 2.8         | 230309 | 57 AC016961   | AL139411 Homo sapi  |
| 25         | 63     | 2.8         | 170387 | 33 AL139411   | AL139411 Homo sapi  |
| 26         | 62.8   | 2.8         | 1080   | 12 MUS9211A   | D38613 Mouse 921-L  |
| 27         | 62.2   | 2.8         | 59686  | 42 AC016204   | AC016204 Homo sapi  |
| 28         | 62     | 2.8         | 16325  | 42 AC015144   | AC015144 Drosophi   |
| 29         | 62     | 2.8         | 63934  | 56 AC010699   | AC010699 Drosophi   |
| 30         | 62     | 2.8         | 106664 | 12 AF130357   | AF130357 Mus muscu  |
| 31         | 62     | 2.8         | 132106 | 56 AC010031   | AC010031 Drosophi   |
| 32         | 61.6   | 2.8         | 315277 | 43 AC018634   | AC018634 Homo sapi  |
| 33         | 61.4   | 2.7         | 68130  | 40 AC004961   | AC004961 Homo sapi  |
| 34         | 61     | 2.7         | 460    | 12 MUSIR3E1   | AC004961 Homo sapi  |
| 35         | 61     | 2.7         | 78467  | 40 AC008078   | M10296 Mouse DNA    |
| 36         | 60.6   | 2.7         | 108409 | 16 AF083424   | AF083424 Aeteline h |
| 37         | 60.6   | 2.7         | 170336 | 32 AL135899   | AL135899 Mus muscu  |
| 38         | 60.6   | 2.7         | 217025 | 33 AL138895   | AL138895 Homo sapi  |
| 39         | 60.4   | 2.7         | 5719   | 12 MNCAT51    | AF051726 Mus muscu  |
| 40         | 60.4   | 2.7         | 90832  | 40 AC003065   | AC003065 Homo sapi  |
| 41         | 60.4   | 2.7         | 113843 | 40 AC005033   | AC005033 , complet  |
| 42         | 60.4   | 2.7         | 234523 | 40 AC006544   | AC006544 , complet  |
| 43         | 60.2   | 2.7         | 110000 | 33 AL137126_2 | Continuation (3 of  |
| 44         | 59.6   | 2.7         | 148700 | 42 AC007492   | AC007492 Homo sapi  |
| 45         | 59.6   | 2.7         | 183570 | 54 AC007906   | AC007906 Homo sapi  |

ALIGNMENTS

RESULT 1  
AF039700 2236 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-38 (NY-CO-38) mRNA, complete cds.  
DEFINITION AF039700  
VERSION AF039700.1 GI:3170199  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
1 (bases 1 to 2236)  
Gordan,J.D., Tureci,O., Sahin,U., Pfrenderschuh,M. and Old,L.J.  
TITLE Characterization of human colon cancer antigens recognized by  
autologous antibodies  
JOURNAL Int. J. Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 2236)  
AUTHORS Scanlan,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1..2236  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene  
1..2236  
/gene="NY-CO-38"  
CDS  
97..2055  
/gene="NY-CO-38"  
/note="PDZ domain protein"  
/codon\_start=1  
/product="antigen NY-CO-38"  
/protein\_id="AAC18049.1"  
/translation="MDRKVAREPRHKVDFLENDAEKDYLDVLRMYHOTMDVAVLVG  
DKLTVNEPSRLPLFDALRPLPLKHQVYEDOLTPRRSKIKRKHREVLINLPRKTY  
RGLEFGCEGLFISHLKGOADSVGLQVDEYLRINGYSISSTTHEEVLINLPRKTY  
SLKVRHIGLIPVKSPPDEPLTWQYVQFVSESGVSGSLGSPENRNRKSKKVFISLVG  
SRGLGSISSGPIQKPIFISHVKGPSLAEVLIGDIVEVNGVDFSNLDHKEAVN  
VKNRSRLSTISYAAGRELFTMDRERLAEARQRELRQBELMOKRLAMESNKIIQEO  
OEHRORKEIKKAOEAENRKYKEMEOIYEERKFKKMEEDMGSEKQDLLPKRTTIA  
EYHVPRLKPKVQGVPELEPADLIDGGEDEGDFPKRYEGEDPYGMYTPEQIMG  
KQVRLRIKKEGSLDALSGVDSPIGKYVYSVYRGAERRGVYKDEYMAINGK  
IYDYTLAEADALQKAMNGDMDLVAVACPKEYDELFLKSKRNOIHALGN  
SELRLPLVNTKPTSLERGHMTTRHMPMDLNSPRLKLPALNQGQIRNSSGHFFE  
GCGGKGAASTRLEDELDKDPDSHSFPLAO"  
BASE COUNT 548 a 588 c 653 g 447 t  
ORIGIN

Query Match 100.0%; Score 2236; DB 40; Length 2236;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 gtcacatgaaccacacgctcgtcctctgtttgatagccattcgcgcgtatccacatg 300  
|||||  
Db 241 gtcacatgaataacccacacgctcgtcctctgtttgatagccattcgcgcgtatccacatg 300  
QY 301 aagcaccaggttgataatgatcagctgaacccccgcgcgtccaggaagctgaagagtg 360  
|||||  
Db 301 AAGCACCAGGTGGAATGATGATCAGCTGACCCCGCGGCTCCAGGAAGCTGAAGAGGTG 360  
QY 361 cgtctgagcgtctgacccccgaagcctcgctcgtatggtgtgtgtggtccgagatt 420  
|||||  
Db 361 Cgtctgagcgtctgacccccgaagcctcgctcgtatggtgtgtgtggtccgagatt 420  
QY 421 ggtctgagcgtctcattcctccacatcaagaagcgttcaggaagcagacgctcgcgc 480  
|||||  
Db 421 GGCTGTGGCTCTTCATCTCCACACCTCATCAAAAGCGGCTGAGGACAGACGCTGGGCTC 480  
QY 481 caagtagggagcagatcgtcgcgataatgatatcatcctcctcctgtatccatgag 540  
|||||  
Db 481 CAGGTAGGGAGCAGATCGTCCGGATCAATGATATTCATCTCTCTTACCCATGAG 540  
QY 541 gaagtaatacactcattcgaaccaagaacatggttccataaagtggacacatcgcgc 600  
|||||  
Db 541 GAGGTATCAACCTCATTCGAACCAAGAAAAGTGTCCATCAAGTGAACACATCGGC 600  
QY 601 ctgaccccgtagaaaagctcctcgtatgaacccctcaacttgacgatgtgatacagtt 660  
|||||  
Db 601 CTGATCCCGTGAAGAAAGCTCTCTGATGAGCCCTCATCTGGCAGATGTGATGATG 660  
QY 661 gtgtcggaaatcgtggggcggtgcgaggaagcgtggtcctccctggaaatcgggaaaacaag 720  
|||||  
Db 661 GTGTGGAATCTGGGGCGGTGCGAGCAGCCTGGGCTCCCTTGAAATGGGAAACAAG 720  
QY 721 gaagaagaagttctcaagcctgtgtaggtcccgagagccttgctgcgaagcattccaag 780  
|||||  
Db 721 GAGAAGAAGTCTTCATCACCTGTGAGGTCTCCGAGGCTTGCTGTCACATTCACAG 780  
QY 781 ggcacacacgaagcctgacatcttatalcagcaatgtaaacctgtcctcgtctcgt 840  
|||||  
Db 781 GGGCCCATCCAGAAAGCTGGCATCTTTATGACGACATGTGAACCTGGCTCCGTCTGCT 840  
QY 841 gaagtggaattgtagatlaaggaggaaccaagattgtcgaagtcaatgctgcactcctaac 900  
|||||  
Db 841 GAGGTGGGATTTGAGATGAGGAGCACAGATGTGGAAGTCAATGGCTCGACTCTCTAAC 900  
QY 901 ctgatacaagaagaagctgtaaatgctgaataaataagccgcgccttgaccatctcatt 960  
|||||  
Db 901 CTGATCACAAGGAGGCTGTAAATGTGCTGAATAATAGCCGACGCTGACCATCTCATTT 960  
QY 961 gtatgctgaactgacgcggagagctgttcattatgacagacgcggagcgttgcaagagcgcg 1020  
|||||  
Db 961 GTAGCTGACAGCTGCGCGGAGAGCTGTTTCATGACAGACCGGAGCGGCTGGCAGAGGCGGG 1020  
QY 1021 cagcgtgaagctgacgcggagagcgttctcattcagaagcgcgtgagatgagttccaac 1080  
|||||  
Db 1021 CAGCGTAGAGCTGACGCGGAGGAGAGCTTTCATGCAAAAGCGGCTGCGATGAGTCCAAC 1080  
QY 1081 aagatctctcagaagcagcagagatgagagcggcgaagaagaaataattgcccagaag 1140  
|||||  
Db 1081 AAGATCTCTCAGGAGGACGACGAGATGAGCGGCAAGGAGAAATAATTGCCCGAGAAG 1140  
QY 1141 gcagcagaagaaatagagatatccggaagagatatgaaacagatg laagagaggaag 1200  
|||||  
Db 1141 GCAGCAGAGGAAATAGAGATACCGGAAGAGATGGAACAGATTTTAGAGAGAGAAAG 1200  
QY 1201 aagtttaagaagaatggaagaagacgtgggtcctaaagaacagctactctgtgcttaa 1260  
|||||  
Db 1201 AAGTTTAAGAAGCAATGGGAAGAGACTGGGCTCAAGGAACAGCTACTCTTGCTTAA 1260  
QY 1261 accataactgctgaggtataaccacagtaacccttcgcaagccaaagtataatacagggagt 1320  
|||||  
Db 1261 ACCATCACTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGATATGATCAGGGAGTG 1320

|            |  |  |                 |
|------------|--|--|-----------------|
| OY         | 1321   | gaacctgagctcgagcccgcaagatgacctcgatgagagcgaggaagcagggagagcag        | 1380            |
| Db         | 1321   | gaacctgagctcgagcccgcaagatgacctcgatgagagcgaggaagcagggagagcag        | 1380            |
| OY         | 1381   | gattccggaatatgtggaagagccttgcagccctactcatctatcttaaccocccagagaatc    | 1440            |
| Db         | 1381   | gatttccggaaatrtatgagaaagcctttgcacccctactctatcttataccccagacagatc    | 1440            |
| OY         | 1441   | atggggaagatgctgcgctcctcaagcatcaagaagaggaatccttagacttgcctg          | 1500            |
| Db         | 1441   | atgggggaagagatgctgcgctcctcaagcatcaagaagaggaatccttagacttgcctg       | 1500            |
| OY         | 1501   | gaagcgcgatgtggaactcccccatctgggaaggtggtcgttctgcctgtatagacgggga      | 1560            |
| Db         | 1501   | gaagcgcgatgtggaactcccccatctgggaaggtggtcgttctgcctgtatagaggggga      | 1560            |
| OY         | 1561   | gctctgaagcgcatgtgtgcatctgtaagaagggacagagatcatggaatcaacgcaag        | 1620            |
| Db         | 1561   | gctctgaagcgcatgtgtgcatctgtaagaagggacagagatcatggaatcaacgcaag        | 1620            |
| OY         | 1621   | atltgacagactaacaccctgcctgctgagctgtaagctgcctctcaagaagagccttgaatcac  | 1680            |
| Db         | 1621   | atlttgacagactaacaccctgcctgctgagctgtaagctgcctctcaagaagagccttgaatcac | 1680            |
| OY         | 1681   | ggcggggagctgtagcatgcacttgtgttcgcgcctctgcctcccaagaaglatagacgtatg    | 1740            |
| Db         | 1681   | ggcggggagctgtagcatgcacttgtgttcgcgcctctgcctcccaagaaglatagacgtatg    | 1740            |
| OY         | 1741   | ctgagccttcttgctgaagttccaaagggaagaacccaattcaagcgttggaaacagtatg      | 1800            |
| Db         | 1741   | ctgagccttcttgctgaagttccaaagggaagaacccaattcaagcgttggaaacagtatg      | 1800            |
| OY         | 1801   | ctccggcgcccaactcctgtaacacaaagccttcgaaaccaagccttgaagagagcgacaatgca  | 1860            |
| Db         | 1801   | ctccggcgcccaactcctgtaacacaaagccttcgaaaccaagccttgaagagagcgacaatgca  | 1860            |
| OY         | 1861   | caacacagatgtagcatccttggtagactgaatctatacaaccaggaatctcaactcccttg     | 1920            |
| Db         | 1861   | caacacagatgtagcatccttggtagactgaatctatacaaccaggaatctcaactcccttg     | 1920            |
| OY         | 1921   | gcccctgaacacgggcccgaataagaacagctcgggcacctttttggaagcgcaatgtbga      | 1980            |
| Db         | 1921   | gcccctgaacacgggcccgaataagaacagctcgggcacctttttggaagcgcaatgtbga      | 1980            |
| OY         | 1981   | ggaagggagcagcgcaacgcttctgggagaagaatctcaagaagatccagactcctcatcctt    | 2040            |
| Db         | 1981   | ggaagggagcagcgcaacgcttctgggagaagaatctcaagaagatccagactcctcatcctt    | 2040            |
| OY         | 2041   | ccctctggcccagtgaaattggtctctcccaagcttggggagactcctccttgaacctaa       | 2100            |
| Db         | 2041   | ccctctggcccagtgaaattggtctctcccaagcttggggagactcctccttgaacctaa       | 2100            |
| OY         | 2101   | taagaaccccaactgagctctctctctcatccctctcctctgcgcccctgtgtctaattgc      | 2160            |
| Db         | 2101   | taagaaccccaactgagctctctctctctcatccctctcctctgcgcccctgtgtctaattgc    | 2160            |
| OY         | 2161   | tgccagatgtgtaactccaacttactctgagctcatatataaataaagatattat            | 2220            |
| Db         | 2161   | tgccagatgtgtaactccaacttactctgagctcatatataaataaagatattat            | 2220            |
| OY         | 2221   | ttccagacttaaaaaa 2236  |                 |
| Db         | 2221   | ttccagacttaaaaaa 2236  |                 |
| RESULT 2   |  |  |                 |
| ABO18687   |  |  |                 |
| LOCUS      | ABO18687   | 2213 bp  | mrna            |
| DEFINITION | Homo sapiens mRNA for autoimmune enteropathy-related antigen AIE-75, complete cds. |  | PRI 02-DEC-1999 |
| ACCESSION  | ABO18687   |  |                 |
| VERSION    | ABO18687.1   | GI:5231270   |                 |
| KEYWORDS   | autoimmune enteropathy-related antigen AIE-75; AIE-75.                             |  |                 |

|                            |   |
|----------------------------|---|
| SOURCE                     | Homo sapiens small intestine cDNA to mRNA.  |
| ORGANISM                   | Homo sapiens<br>Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;<br>Eutheria; Primates; Catarrhini; Hominiidae; Homo.  |
| REFERENCE                  | 1 (sites)<br>Kobayashi, I., Imamura, K., Kubota, M., Ichikawa, S., Yamada, M.,<br>Tonoki, H., Okano, M., Storch, W.F., Morituchi, T., Sakiyama, Y. and<br>Kobayashi, K.<br>Identification of an autoimmune enteropathy-related 75-kilodalton<br>antigen   |
| TITLE                      | Gastroenterology 117 (4), 823-830 (1999)  |
| JOURNAL                    | 99431904  |
| MEDLINE                    | 2 (bases 1 to 2213)   |
| REFERENCE                  | Kobayashi, I.   |
| AUTHORS                    | Direct Submission<br>Submitted (14-OCT-1998) to the DDBJ/EMBL/Genbank databases. Ichiro<br>Kobayashi, Hokkaido University School of Medicine, Department of<br>Pediatrics, North-15, Kita-ku, Sapporo, Hokkaido 060-8638,<br>Japan (E-mail: ichikoba@med.hokudai.ac.jp, Tel:011-716-1161,<br>Fax:011-736-9267)  |
| FEATURES                   | location/Qualifiers   |
| source                     | 1..2213<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/tissue_type="small intestine"<br>90..1748<br>/codon_start=1<br>/product="autoimmune enteropathy-related antigen AIE-75"<br>/protein_id="BAAB1740.1"<br>/db_xref="GI:5231271"<br>/translation="MDRRVAKEFRHKVDFLENDAEKDYLVDLRMTHTQMDVAIVLG<br>DLKLVINEPSRLPLFDAIRPLPKHVEVDQLPPRSRKIKERLDRLHPBEGLSV<br>SGKEFGCGLFISHLIKGGQADSVGLDVGDEIYINISLSSCHEEVNLIPIRTKV<br>RIKRGHGLIPKVSFDPDELTVDDVFSSGGSVRSLSPGRKKRKKVFTSLVG<br>SRGISCSISGPLOKPGIFLSITSHVKGSLSAEVLGETDOIYENVNGNSLNKLEANV<br>VLKSRSILTITIVAAGRELPMDRELRAKROELROEILMOKRLAMSNKLQGO<br>OEMBRQRKREITAOKRAAENERYKRMEQIVIEEEKFKQWDEMGSEQLLPPTTA<br>EVNHVPFNKPYYDGVPELEPADDDLGTEGEODPFRKYEEGFDSMTPTQING<br>KDVLLRIKPKKESIDLALEGVDSDPIKVVVASVAYERGAERHGIVKGDIEIMALINK<br>IVDTVTAAEAFAALOKMANOGCDWIDLVAVCPRKYEDELTFP" |
| CDS                        | polyA_site<br>2213<br>/note="8 a nucleotides"<br>BASE COUNT 539 a 587 c 650 g 437 t<br>ORIGIN   |
| Query Match                | 95.7%; Score 2139.4; DB 9; Length 2213;   |
| Best Local Similarity      | 99.4%; Pred. No. 0;   |
| Matches 2168; Conservative | 0; Mismatches 11; Indels 2; Gaps 2;   |
| Oy                         | 8 cggtcgagtcgcgactctttccagaacctcctggcaacccgggaagaacgagtctg 67<br>   |
| Db                         | 1 CGGTGCGGTCGCGGCTCTTTCCACTCTCTGGCACCGGCAACCGAAGCGGTGCT 60<br>  |
| Oy                         | 68 gcaacgacgacgttgagactggcccaccalcattggaacgaaagtggccccagaatccggc 127<br>  |
| Db                         | 61 GCAACGACGACACTGGACTGGCCACGACGATGAGCAAAGTGCGCCGAGATTCCGGC 120<br>   |
| Oy                         | 128 ataaggtgatcttcttgattgaaatatgatgcagagaagaagactactcatgatgtgtcgc 187<br>   |
| Db                         | 121 ATAAAGGTGGATTTCCTGATTGATAAATGATGCAGAAGAACACTATCTCATGATGTGCTGC 180<br>   |
| Oy                         | 188 gaatttacccaagaacccatlgacgtggccgcgctgcgttgaggaaaccttgaaagctggatcatca 247<br>   |
| Db                         | 181 GAATTTACCACCAACCATGAGACGTGGCCGCTGCTGTGTGGAGAACTTGAAGCTGGTCAATCA 240<br>   |
| Oy                         | 248 atgaaccccaagccgctgtcctctgttttgtatgtccattgcgcgcgttgatcccaactgaaagcacc 307<br>  |
| Db                         | 241 ATGAACCCACGCGCTGCTGCTGTGTTGATGCCATTTCGGCCGCTGATCCCAAGAACACCC 300<br>  |
| Oy                         | 308 aggtgatgaatgatcatgactgaccccccgagcgtccaggaaagctggaagagtgcgtctgg 367<br>  |
| Db                         | 301 AGGTGAATATCATATAGCTGACCCCCCGCGCGCTCCAGGAAGCTGAAAGAGGTGCGTCTGG 360<br>   |

|    |      |  |      |
|----|------|--|------|
| Qy | 368  | acgcctgcaccccccgaagggcctcgccctgaagctgagtgctgagctggaatttgctg      | 427  |
| Db | 361  | ACGCCTGCACCCCGAAGGGCTCGCCCTGAAGTGTGTGTGTGGCTGGAAGTTGGCTGTG       | 420  |
| Qy | 428  | ggcctctcatctcccacatcatcaaaagggctgcagggcagacagcgctcgggtccaggtag   | 487  |
| Db | 421  | GGCTCTCATCTCCACCTCATCTCAAAAGGGGGTCCAGGCACAGCGTGGGGCTCCAGGTAG     | 480  |
| Qy | 488  | ggagcagatcgctccggatcaatgatgatccaatcctctctgtaccgaaggagatca        | 547  |
| Db | 481  | GGGAGCAGATCGTCCGGATCAATGGATTTCCATCTCCCTGTACCCATGAGGAGGTCA        | 540  |
| Qy | 548  | tcaaccctattgaaaccaaaaacgcttccatcaaaatggagacaactggcctgtacc        | 607  |
| Db | 541  | TCAACCTCATTCGAAACCAAAAACGTGTCCATCAATGAATGAGACATCGGCTGATCC        | 600  |
| Qy | 608  | ccgtgaaaagctcctcctgataagccccactctggcagatctggatcaattgtgtcgg       | 667  |
| Db | 601  | CCGTGAAAAGCTCCCTCGATGAGACCCCTCACTTGGCAGTATGTGATCAGTTTGTGTGG      | 660  |
| Qy | 668  | aactggggcgctgcgagagcagccctgggctcccttggaaatcgggaaaaacaagagaaga    | 727  |
| Db | 661  | AATCTGGGGCGTCCGAGCAGCCTGGGCTCCCTGGGAATGGGAAAAACAAGGAGAAGA        | 720  |
| Qy | 728  | agcgtctcatcaacgttggatggtctcccaagggccttggctcgagaaatctcagggcccca   | 787  |
| Db | 721  | AGGCTCTCATCAAGCTGTGTAGGCTCCCGAGGCTTGTGGCTGCAAGATTTCCAGCGGCCCA    | 780  |
| Qy | 788  | tccagaagccttgcatacttattatcaagccatgttbaaaccttgcctccctgtctctgagttg | 847  |
| Db | 781  | TCCAGAAGCCTGGCATCTTATCAGCCATGTGAAACCTGGCTCCTGCTGTGTGAGGTGG       | 840  |
| Qy | 848  | gattggagataaggggagaccagattgttcgaagtcaatggcgttgactctcttaactgtgac  | 907  |
| Db | 841  | GATTGGAGATAAGGGGAGACCAATGTTGTGAAGTCAATGGCTCCACTTCTCTAACTGGATTC   | 900  |
| Qy | 908  | acaaaggagcctgttaaaatgtctgtaaaaatagcccgagcccttgacatctccatttaagt   | 967  |
| Db | 901  | ACAAAGGAGCCTGTAAATGTCTCTBAAGATGACCGCAGCTTGACACTTCTCCATTGTAGCTG   | 960  |
| Qy | 968  | cagcttgcgcggagcgtgttcatatgacagaccggagcggtgtgcagagggcgcgagcgt     | 1027 |
| Db | 961  | CAGCTGCGCGGAGCTGTTCATGTACAGACCGGAGGCGCTGSCAGAGGCGGGCAGCGGTG      | 1020 |
| Qy | 1028 | agctgcagcgagcagagcttctcaatgcagaagcgcgtgtgcgatgttgatctcaacaagatcc | 1087 |
| Db | 1021 | AGCTGCAGCGCGAGGAGCTTCTCATCTCAAGACGGGCTGCGATGTGAGTCCAAACAAGATCC   | 1080 |
| Qy | 1088 | tccagggagcagcgaggaattggagccggcaaaaggaagaaanaatttcccaagagcaagag   | 1147 |
| Db | 1081 | TCCTGGGAGCAGCGAGATGTGACGGCCAAAGGAGAAATAATTTCCACGAAGGAGGAG        | 1140 |
| Qy | 1148 | aggaaaaaatgagagatcacccggaagaagtgaagaacagatttlaagagaggaagaagattta | 1207 |
| Db | 1141 | AGGAAAAATGAGAGATACCGGAAGGAGATGGAACAGATTGTAGAGGAGGAAGAAAGTTTA     | 1200 |
| Qy | 1208 | agaagcaatvggaaagaagacttgggtctcaaaagaaacagctactcttgccttaaacatca   | 1267 |
| Db | 1201 | AGAAAGCAATGGGAAAGAGACTGTGGGCTCAAAAGAACAGCTACTCTTTCGCTTAAACATCA   | 1260 |
| Qy | 1268 | ctgtctgggtgtaacaccagtaaccccttcgaaagccaaagatgatataaggaattgaaactg  | 1327 |
| Db | 1261 | CTGCTGAGGTACACCCAGTACCCCTTCGCAAGCCAAAGATGATATAGGAGAGTGGAACTTG    | 1320 |
| Qy | 1328 | agctcgaagcccgagatgtacctgtgatgtgagacgagaggaaggaagagagagatcttc     | 1387 |
| Db | 1321 | AGCTCGAGCCCGAGATGAGACTGTGATGTGAGGCGACGAGAGGACGAGGAAGCAGGATTTCC   | 1380 |
| Qy | 1388 | ggaaatatgagaaagacttggaccctactctatgttcaacccagagcagatcatyggga      | 1447 |
| Db | 1381 | GGAATATTGAGGAAGGCTTTGACCCCTACTCTATGTTCACCCCAAGCAGATCATGTGGGA     | 1440 |

|            |  |   |                |
|------------|--|---|----------------|
| QY         | 1448   | aggatgtccggctctctacgcgtacaaagaagagatccttaaaccttgcccttggaagcg    | 1507           |
| Db         | 1441   | AGGATGTCCGGCTCTCTACGCATCAGAAGAGAGGATCTTAGACCTGGCCCTGGAAAGCG     | 1500           |
| QY         | 1508   | gtctggaactcccccacttgggaagtgctgtcttctctgtatgaagcggagactgtcgt     | 1567           |
| Db         | 1501   | GTGTGAGATCTCCCATTTGGGAAAGGTGTCGTTCTGCTGTGATGACGGGAGCGTCTGTG     | 1560           |
| QY         | 1568   | agcgagcatggtgagtttgtgaaggggagcagatctatgtcaatcacaagcgaatltgtga   | 1627           |
| Db         | 1561   | AGCGGCAGGTGGGATTTGTGAAGAGGGAGCAGATCATGGCAATCAACGGCAAGATTGTGA    | 1620           |
| QY         | 1628   | cagaetacacccctgtgcgtgagctgcagctcctctgcagaagccttgaatctcagggcggg  | 1687           |
| Db         | 1621   | CAGACTACACCCCTGGCTGTGAGGCTGAGGCTCCTCGCAGAAAGCGCTGTGATTCAGGGCGGG | 1680           |
| QY         | 1688   | actgtagtcagacctgtgtgttgcgctctgcctcccccagaagagatgaagtagtcagctgac | 1747           |
| Db         | 1681   | ACTGGAATCGACCTGTGTGGTTGCCGTCTGCCCCCAAGAGATGATGCATGATGACCTGACT   | 1740           |
| QY         | 1748   | tctgtctgaatccaaaaagggaaccaaattacgcgcttagaagaacagttagctccggc     | 1807           |
| Db         | 1741   | TCCT-CTGAATCTCCAAAGGGGAAACCAAAATTCAC-CTTTAGGAATACGTGAGCTCCGG    | 1798           |
| QY         | 1808   | ccacactcgttgaacacaaagcctctgcagccagcctltgagagagccacatgacacacaca  | 1867           |
| Db         | 1799   | CCCACTCTGTGACACAAAGCGTGTGATCAGCTTGAGAGAGGCCACACTACACACACA       | 1858           |
| QY         | 1868   | gatgtagcatccttggagaccgtgaatctatcacccagagatctcaactcccttggccctga  | 1927           |
| Db         | 1859   | GATGGCATCTCTTGGACCTGTGAATCTATCACCCAGGAATCTCAAACTCCCTTTGGCCCTGA  | 1918           |
| QY         | 1928   | accagggccaataaaggaacacagctctcggccacttlttgaagccaatltgtaggaaag    | 1987           |
| Db         | 1919   | ACCAAGGCCCAATFATGAAGAACACGCTGGGGCCACTCTTGTAAGGCCCAAGCTGGAGAAAG  | 1978           |
| QY         | 1988   | gagcagccagccglttgggaggaagatctcaagatccagactctaatcttctctcttgg     | 2047           |
| Db         | 1979   | GAGCAGCCAGCCATTTGGGAAAGATCTCAAGATCCAGACTCTCTTCTTCTCTCTGG        | 2038           |
| QY         | 2048   | ccagatgaattgtctctcccaagccttggggagaccccttcccttgaacctataaagacc    | 2107           |
| Db         | 2039   | CCCACTGAATTTGGTCTCTCTCCACACTCTGGGGGACTCTCTTCTTGAAACCTTATTAAGCC  | 2098           |
| QY         | 2108   | ccactggaatctctctctctccatccctccctcctctgcctctcgtcttaatltgtctcag   | 2167           |
| Db         | 2099   | CCACTGGAGTCTCTCTCTCTGTCATACCTCTCCCTGCGCTGCTGCTATATGTCTCCAGG     | 2158           |
| QY         | 2168   | attgtcaatccaactctact  | 2188           |
| Db         | 2159   | ATTGTCACTCCCAACACTTACT  | 2179           |
| RESULT 3   |  |   |                |
| LOCUS      | AF039699   |   |                |
| DEFINITION | Homo sapiens   | 2162 bp   | mRNA           |
| ACCESSION  | AF039699   |   |                |
| VERSION    | AF039699.1   |   |                |
| KEYWORDS   |  |   |                |
| SOURCE     | human.   |   |                |
| ORGANISM   | Homo sapiens   |   |                |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  |   |                |
| AUTHORS    | 1 (bases 1 to 2162)<br>Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E., Jordan,J.D., Turci,O., Sahin,U., Pfrendeschuh,M. and Old,L.J. |   |                |
| TITLE      | Characterization of human colon cancer antigens recognized by autologous antibodies  |   |                |
| JOURNAL    | Int. J. Cancer   | 76 (5),   | 652-658 (1998) |
| REFERENCE  | 98272252   |   |                |
| AUTHORS    | 2 (bases 1 to 2162)<br>Scanlan,M.J.  |   |                |

TITLE Direct Submission  
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021, USA

FEATURES  
source  
1. .2162  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colon cancer metastasis to lung"  
1. .2162  
/gene="NY-CO-38"  
97. .1308  
/gene="NY-CO-38"  
/note="PDZ domain protein; NY-CO-38 antigen splice variant"  
/codon\_start=1  
/product="antigen NY-CO-38"  
/protein\_id="AAC18048.1"  
/db\_xref="gi:3170198"  
/translation="MDRKVAREPRHKVDLEIENAKRDIYDLRMVHQTMVAIVG  
DKLIVNPSRLPIFDATRPILPKHVEYQQLPRRSRKKEVRLDLHPBGLSLV  
RGLEFGGLTSHLIRGGQADSVGLQVDEIVRINGISSTHEEVLNLRIRKTV  
SIKVRHIGLIPKSSPDEPLTMQYVDQVSSGVRSLGSPGNRENEKRVFTSLVG  
SRGLGCSISGPIQKPGIFISHVKGSLSAEVLGIEQIYEANGVDFSNLDHKEAV  
VLKNSRSLTISIVAAGRELEMTDRELRLEAERORELOELMOKRLMESNKILIOE  
QEMERORKEITAOAAEENERYRKMEQIIVEBEKFKQWEDMSKEQLLPTIITA  
EVHPLEPKPKP"

BASE COUNT 528 a 573 c 622 g 439 t

ORIGIN

Query Match 92.9%; Score 2078; DB 40; Length 2162;  
Best Local Similarity 96.7%; Pred. No. 0; Mismatches 0; Indels 74; Gaps 1;  
Matches 2162; Conservative 0; Mismatches 0; Indels 74; Gaps 1;

QY 1 cctggcccgctcgcgctccttccagctcctgcaagccggccagcaagaaac 60  
DB 1 cctggcccgctcgcgctccttccagctcctgcaagccggccagcaagaaac 60  
QY 61 gggctcgtcaacgacgacgtgacccgtgcccagcaatgacgaagtgccgagaa 120  
DB 61 gggctcgtcaacgacgacgtgacccgtgcccagcaatgacgaagtgccgagaa 120  
QY 121 ttccggcaataagtgatgttcttgatgaataatgacgaagaaagacatctatgt 180  
DB 121 ttccggcaataagtgatgttcttgatgaataatgacgaagaaagacatctatgt 180  
QY 181 gtcgtcgaatgtacacacacacatgacgtgacgtgacgtgacgtgacgtgac 240  
DB 181 gtcgtcgaatgtacacacacacatgacgtgacgtgacgtgacgtgacgtgac 240  
QY 241 gtcacaaatgaac 300  
DB 241 gtcacaaatgaac 300  
QY 301 aagcccaagtgagaaatgatgatgatgatgatgatgatgatgatgatgatgat 360  
DB 301 aagcccaagtgagaaatgatgatgatgatgatgatgatgatgatgatgatgat 360  
QY 361 cgtccggaacgtctgacacacacacacacacacacacacacacacacacacac 420  
DB 361 cgtccggaacgtctgacacacacacacacacacacacacacacacacacacac 420  
QY 421 ggcgttgagcctcttcatctccacacacacacacacacacacacacacacac 480  
DB 421 ggcgttgagcctcttcatctccacacacacacacacacacacacacacacac 480  
QY 481 caggtgaaggaagaaatgctcggatcaatgatgatgatgatgatgatgatgat 540  
DB 481 caggtgaaggaagaaatgctcggatcaatgatgatgatgatgatgatgatgat 540  
QY 541 gaggctacaacatcatctgaacacaaagaacatgtgtcatcaaatgtgacacat 600  
DB 541 gaggctacaacatcatctgaacacaaagaacatgtgtcatcaaatgtgacacat 600

DB 541 GAGGTGATCATCAACCTCATTCGAAACCAAGAAACTGTGTCCATCAAAAGTGACACATCGCG 600  
QY cgtatcccccgtgaataaagctctctctgatgagccccctactgtgcagfatgtgatcaatt 660  
DB 601 cgtatcccccgtgaataaagctctctctgatgagccccctactgtgcagfatgtgatcaatt 660  
QY 661 gtcgcgaatctgagggcgtgaggaagcagcctggcctcccttggaatgggaaaaaag 720  
DB 661 gtcgcgaatctgagggcgtgaggaagcagcctggcctcccttggaatgggaaaaaag 720  
QY 721 gagaagaaggtctcatcagcctgtaggctcccgagccttgctgtagcagcatllccagc 780  
DB 721 gagaagaaggtctcatcagcctgtaggctcccgagccttgctgtagcagcatllccagc 780  
QY 781 gggcccaatcagaagcctgtagccttcatcagccatgtgaaacgtgctccctgtctgt 840  
DB 781 gggcccaatcagaagcctgtagccttcatcagccatgtgaaacgtgctccctgtctgt 840  
QY 841 ggggtgagatggagaaaggaagcagatgtcgaaatcaatgagcgtgagcttctctaac 900  
DB 841 ggggtgagatggagaaaggaagcagatgtcgaaatcaatgagcgtgagcttctctaac 900  
QY 901 cctgatacaagaagagctgttaaatgtctgaaataatagccgagcctgacacatccat 960  
DB 901 cctgatacaagaagagctgttaaatgtctgaaataatagccgagcctgacacatccat 960  
QY 961 gtagctgacagctgagccgagagctgttcatgacagacccgagcgtgacagagcgcg 1020  
DB 961 gtagctgacagctgagccgagagctgttcatgacagacccgagcgtgacagagcgcg 1020  
QY 1021 cagcgtgagcctgcaagcagagagcttctcatgcaagcagcgtgagcagatcccaac 1080  
DB 1021 cagcgtgagcctgcaagcagagagcttctcatgcaagcagcgtgagcagatcccaac 1080  
QY 1081 aagatctccagagagcagagatgtagcgcgcaagaagagaagaataatgtcccaag 1140  
DB 1081 aagatctccagagagcagagatgtagcgcgcaagaagagaagaataatgtcccaag 1140  
QY 1141 gcaagcagaggaataatgagagataccggaagagatggaacagattgtagaagaagaag 1200  
DB 1141 gcaagcagaggaataatgagagataccggaagagatggaacagattgtagaagaagaag 1200  
QY 1201 aagtttaagaagaacatggaagaagaacatgagcgtgcaagaagaacagcttctgca 1260  
DB 1201 aagtttaagaagaacatggaagaagaacatgagcgtgcaagaagaacagcttctgca 1260  
QY 1261 accatcaactgctgaggtacacacacacacacacacacacacacacacacacacac 1320  
DB 1261 accatcaactgctgaggtacacacacacacacacacacacacacacacacacacac 1320  
QY 1321 gaaactgagctgagccgagagatgacctgtagaggaaggaaggaaggaaggaag 1380  
DB 1321 gaaactgagctgagccgagagatgacctgtagaggaaggaaggaaggaaggaag 1380  
QY 1381 gatttcggaataatgaggaagccttgaacccctactactaattgtcaaccccaagcagatc 1440  
DB 1381 gatttcggaataatgaggaagccttgaacccctactactaattgtcaaccccaagcagatc 1440  
QY 1441 atcggaagaagatgtccggtctctcagcatalcaagaagaaggaaggaaggaaggaag 1500  
DB 1441 atcggaagaagatgtccggtctctcagcatalcaagaagaaggaaggaaggaaggaag 1500  
QY 1501 gaaagcgtgtgagactcccatctggaaggtgtgtcgttctgctgtgtatgagcggtga 1560  
DB 1501 gaaagcgtgtgagactcccatctggaaggtgtgtcgttctgctgtgtatgagcggtga 1560  
QY 1561 gcttcgtgagcgtgagctgtgagcattgtgaaaggaaggaaggaaggaaggaaggaag 1620  
DB 1561 gcttcgtgagcgtgagctgtgagcattgtgaaaggaaggaaggaaggaaggaaggaag 1620  
QY 1621 atgtgacagactacacccctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 1680  
DB 1621 atgtgacagactacacccctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 1680  
QY 1681 atgtgacagactacacccctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 1740  
DB 1681 atgtgacagactacacccctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgtgag 1740

|            |  |   |             |
|------------|--|---|-------------|
| QY         | 1681   | ggcgggagctgatgcacaccttggtgtccgtctgtcccccaaaaggagatgaagatgag     | 1740        |
| Db         | 1607   | GGCGGGGAGCTGTGATCGACCTTGTTGGTTCGCCCTCTGCCCCCAAAAGAGTAGTCAGATGAG | 1666        |
| QY         | 1741   | ctgacctcttgcctcfaaatgccaaaagggaaccctaattcaocggttaagaacaagttag   | 1800        |
| Db         | 1667   | CTGACCTTCTTGCTGAATCCAAAAGGGGANAACCATAATCACGGGTTAGAAGAACAGGAG    | 1726        |
| QY         | 1801   | ctccggcccacaactcgtaaacaagaaccttcggaccaagcctttgagagggccaatgaca   | 1860        |
| Db         | 1727   | CTCCGGCCCCACACTGTGTAACAACAAGGCCTCGAGACAGCCTTGAGAAGGCCACATGACA   | 1786        |
| QY         | 1861   | cacaccagatgcatcctcttggagcctgaatatcatcacccaggaatccaactcccttgg    | 1920        |
| Db         | 1787   | CACACCAATGGCATCTCTTGGAGCACTGAAATATCACCCAGGAATTCCAAATCCCTTTG     | 1846        |
| QY         | 1921   | gccctgaacccaaggccaataaagaacagcctcggggcaactttttgaaggccaatgtga    | 1980        |
| Db         | 1847   | GCCCTGAACCAAGGCCACATATAGAACAGCTCGGGCCAATTYYYGAAGGCCATGTGA       | 1906        |
| QY         | 1981   | ggaagaaggagcagcagcagcglttgggaaagatccaagatccaagatccaacttcatt     | 2040        |
| Db         | 1907   | GGAAGGAGAGCAGCAGCGCTTGGAGAAAGATCATCAAGATCOAAGCTCATCTCCTT        | 1966        |
| QY         | 2041   | cctctggcccagtgatatttgtctctccagccttggggagctccttcttgaacctta       | 2100        |
| Db         | 1967   | CCTCTGGCCCAGATGATTTGGTCTCTCCACGCTTTGGGGGACCTCTCTCTTGAACCTTA     | 2026        |
| QY         | 2101   | taagaccccatctgagctctctctctccaaacctctccctcctcctcctcgtcattaigc    | 2160        |
| Db         | 2027   | TAAAGCCCCACTGAGAGTCTCTCTCTCCATCCCTCTCTGCTCTGCTCTGCTTAATTGC      | 2086        |
| QY         | 2161   | tgcagagatgtcatcctccaaccttactcttgagctcataataaataacagattatt       | 2220        |
| Db         | 2087   | TGCCAGGATTTGCTACCTCCAAACCTTACTGTGAGCTCATTAATAAATAACAGATTATT     | 2146        |
| QY         | 2221   | ttccagcttaaaaaa   | 2236        |
| Db         | 2147   | TTCCAGCTTAAAAAA   | 2162        |
| RESULT     | 4  |   |             |
| AB006955   | LOCUS  |   |             |
| DEFINITION | AB006955   | 2283 bp mRNA  | PRI         |
| ACCESSION  | AB006955   | Homo sapiens mRNA for AIE-75, complete cds.                     | 02-DEC-1999 |
| VERSION    | AB006955.1   | GI:5152287  |             |
| KEYWORDS   | AIE-75; aie-75.  |   |             |
| SOURCE     | Homo sapiens   | small intestine cDNA to mRNA.                                   |             |
| ORGANISM   | Homo sapiens   |   |             |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;                  |   |             |
| AUTHORS    | Eutheria; Primates; Catarrhini; Hominiidae; Homo.                    |   |             |
|            | 1 (sites)  |   |             |
|            | Kobayashi, I., Imamura, K., Kubota, M., Ishikawa, S., Yamada, M.,    |   |             |
|            | Fonoki, H., Okano, M., Storch, W.F., Moriwuchi, T., Sakiyama, Y. and |   |             |
|            | Kobayashi, K.  |   |             |
| TITLE      | Identification of an autoimmune enteropathy-related 75-kilodalton    |   |             |
|            | antigen  |   |             |
| JOURNAL    | Gastroenterology   | 117 (4), 823-830 (1999)   |             |
| MEDLINE    | 99431904   |   |             |
| REFERENCE  | 2 (bases 1 to 2283)  |   |             |
| AUTHORS    | Kobayashi, I.  |   |             |
| TITLE      | Direct Submission  |   |             |
| JOURNAL    | Submitted (31-AUG-1997) to the DDBJ/EMBL/GenBank databases. Ichiro   |   |             |
|            | Kobayashi, Hokkaido University School of Medicine, Department of     |   |             |
|            | Pediatrics, North-15, West-7, Kita-ku, Sapporo, Hokkaido 060-8638,   |   |             |
|            | Japan (E-mail: ichikoba@med.hokudai.ac.jp, Tel:011-716-1161,         |   |             |
|            | Fax:011-736-9267)  |   |             |
| FEATURES   | Location/Qualifiers  |   |             |
| SOURCE     | I. .2283   |   |             |
|            | /organism="Homo sapiens"   |   |             |
|            | /db_xref="taxon:9606"  |   |             |

| gene | /tissue_type="small intestine"  |
|------|---|
| 96   | 246..1611   |
| 97   | /gene="aie-75"  |
| 98   | 246..1611   |
| 99   | /gene="aie-75"  |
| 100  | /codon_start=1  |
| 101  | /product="Aie-75"   |
| 102  | /protein_id="BA81739.1"   |
| 103  | /db_xref="GI:5152288"   |
| 104  | /translation="MYHQTMDVAVLVGDLKLVINEPSRLPLFDAIRPLILPKHQVEYI<br>OLTPRSRLKEVRLDRDLHPREGLTGSVSGEGFEGFLTSHLRKGGADSVGLGVDSD<br>IVRINGYSISCTHEEVNLIRTKTVAIVRHGILPVKSSDPDLTMQVDDPVS<br>SGVAGSGISGSPGNRENKREKYPILSVGRGLGCSISSPQIKPPIFISHVPRGSLSA<br>VLEIGDQIVVEGNVDFSNLHKREAVNYLKSRSRTLSIVAAGSELEWTRERLTEL<br>RORELQROELMLQKRLAMESSNKILQEDMEMEORRKLTAQNAEENRYRREMEIVAE<br>EERKQKQMEEDMGSEKQOLLPLKTTAAVHVPLRPKYDQGEVELEPADLDGDTG<br>EOGEDDFKREVEGFDYPSMPTPEQIIMGADVRLRLKRGSLDLALEGVDVSPICKVVV<br>SAVYRGAAERHGRGIVKDELMIAINGKIVYDTITLAEALALQKAMNGSDMIDLVAVA<br>CPPREYDEDELTFP" |
| 105  | BASE COUNT 549 a 594 c 670 g 470 t  |
| 106  | ORIGIN  |
| 107  | Query Match 89.5%; Score 2001.6; DB 9; Length 2283;   |
| 108  | Best Local Similarity 98.2%; Pred. No. 0;   |
| 109  | Matches 2068; Conservative 0; Mismatches 29; Indels 9; Gaps 4.  |
| 110  | 131 aggtgatttccttgattggaataatgtagagaagaagactatctcatatgctgtcggaa 190   |
| 111  |   |
| 112  | 187 aggtgattttcttgattggaataatgtagagaagaagactatctcatgctgacga 246   |
| 113  |   |
| 114  | 191 tgtaccaccagaaccatggaagctgagcgcgtgctcgttggagaacctgaagctgcatcaatg 250   |
| 115  |   |
| 116  | 247 tctacacacagacatgagacgtggccgctgctgtggagaacctgaagctgcatcaatg 306  |
| 117  |   |
| 118  | 251 aaccagccgtctgcctctgttcttgatgcattcgccgcgtgatcccaatgaagcaaccag 310  |
| 119  |   |
| 120  | 307 AACCCAGCCGCTGCTCTGTTGATGATGACATTCGCGCCGCTGATCCACATGAAGCACAG 366   |
| 121  |   |
| 122  | 311 tggaaatgatagctatgaccccccggcgctccagaagaatggaagagtgcttggacc 370   |
| 123  |   |
| 124  | 367 tggaaatgatagatagctgacccccggcgctccagaagaatggaagagtgcttggacc 426  |
| 125  |   |
| 126  | 371 gctctgaccccgaaagccctcgagctgagtgctgctgagtgcccttgaagtttgctgtg 430   |
| 127  |   |
| 128  | 427 gctctgaccccgaaagccctcgagctgagtgctgctgagtgcccttgaagtttgctgtg 486   |
| 129  |   |
| 130  | 431 tcttcatctccacatcaaaagcggtcagagcaacgctcggtctcaagtaagg 490  |
| 131  |   |
| 132  | 487 tcttcatctccacatcaaaagcggtcagagcaacgctcggtctcaagtaagg 546  |
| 133  |   |
| 134  | 491 acgagaatgcctcgagtaagatgatatctactctctctgtaaccataagaggtcatca 550  |
| 135  |   |
| 136  | 547 ACGAATGCTGCGGACATGATGATTCATCTCTCTGTAACCATGAGGAGGTCATCA 606  |
| 137  |   |
| 138  | 551 acctcatcgaaccagaanaactggtcatcaaaagttagacacatcgccgtgaccccg 610   |
| 139  |   |
| 140  | 607 ACGTCAATTCGAACCAAGAAACTGTGTCTCATCAAGTAGACACATCGGCTCATCCCG 666   |
| 141  |   |
| 142  | 611 tgaagaagctctctctgtagagccctcacttggcagtaatgtagatcagtttgttcgaat 670  |
| 143  |   |
| 144  | 667 TGAAGAAGCTCTCTCTGATGAGCCCTCACTTGGCAGTATGTGGATCATGTTGTCGGAAT 726   |
| 145  |   |
| 146  | 671 ctgggggcgtctcgagagcagccttggtcctcccttggaatcgggaaacaagaagaag 730  |
| 147  |   |
| 148  | 727 CTGGGGGCGTCTCGAGGCGCTCGGCGCTCCCTCGGAAATCGGGAACCAAGAGGAAGG 786   |
| 149  |   |
| 150  | 731 tcttcatcagcttgtagctccccaagccttggtgtagagatttcaggcgccccatcc 790   |
| 151  |   |
| 152  | 787 TCTTCATCAAGCTGTGATGAGCTTCGAGGCGCTTGGCTGACGATTTCCACGCGCCCATATC 846   |
| 153  |   |
| 154  | 791 agaaagcctgagcatcttctcagcagatggaacacctgagctccctgtctcgtagtgagat 850   |



Db 847 AGAAGCCTGGCATCTTATCATCAGCCATGTGAACCTGGCTCCCTGTGCTGAGTGGAT 906  
 QY 851 tggagataaggagaccagatttgtgaagtcaatggcgtcgactctctaacttgatcaca 910  
 Db 907 TGGAGATAGGGGAGCACAAATGTGCAAGTCAATGCGTGCATTTCTTAACCTGGATCACA 966  
 QY 911 aggaagctgtaaatgtctgtaaaaaatagccgcagcctgacacatctgcattgttagctcag 970  
 Db 967 AGGAGGCTGTAAATGTCTTGAAGATGAGCCGACCTGACCATCTTCCATTGTACTTCAG 1026  
 QY 971 ctggccggagagctgttcaatcacagaccgggagcggcttgacagagcgccggcagctgac 1030  
 Db 1027 CTGGCCGGAGCTGTTGATGACAGACCGGGAGCGGCTGACAGAGGCGCGGAGGCTGAGC 1086  
 QY 1031 tgcagcgagagagctctcatgacagagcgctgacatgagatccacaagaatctcc 1090  
 Db 1087 TGGAGCGGCGAGGAGCTTCTATGAGAGCGGCTGCGATGAGATCCAAAGATCTCTCC 1146  
 QY 1091 aggaagcagagagagctgagcggcgaagagagaaatattgcccagaagcgagcagagg 1150  
 Db 1147 AGGAGCGAGGAGATGAGCGCGCAAGAGAGAAATTTGCCAGAGAGCGACAGAGG 1206  
 QY 1151 aaaaatgagagataccggaagagagatggaacagattgtagaagagaaagtttaaga 1210  
 Db 1207 AAATATGAGATATACCGGAGAGAGATGGAACAGATTGTAGAGAGAGAGAGATTTAAGA 1266  
 QY 1211 agcaatggaagaagagacttgggtctcaagaaacagactactctgtcttaaaacatactg 1270  
 Db 1267 AGCAATGAGGAGAAAGACTGGGGCTCAAAAGAGAACAGCTACCTTCCCTAAACCATCACTG 1326  
 QY 1271 ctgaaggaacccagtaaccccttcgcaagcgaagatgatagtcaggaggttgaacattagg 1330  
 Db 1327 CTGAGGTACACCGGATACCCCTTGCGAAGCCAAAGATGATGACGGGATGGAACTGAGC 1386  
 QY 1331 tcgaagcccgagatgacactgtagatgagcagcagagagcagggagaaacagattccgga 1390  
 Db 1387 TCGAGCCCGGAGATGACTGATGAGAGGCAAGGAGAGAGAGAGAGATTTCCGGA 1446  
 QY 1391 aatatgaggaaggctttagccctcctatgttcaaccacagagacatcatggyggaag 1450  
 Db 1447 AAATATGAGGAGGCTTGTGACCCCTCACTCTATGTTCAACCCAGAGCATATGAGGGAAG 1506  
 QY 1451 atgtccggctcctaagataaagaagagagatctctaagacctggccctggaagcggtg 1510  
 Db 1507 ATGTCCGGCTCTACGATGAAGAGAGGAGATCTTAGACCTGGCCCTGGAAGCGGCTG 1566  
 QY 1511 tggactcccccatgtgaaaggtgtcgttctcgtgtgtatgagcgggagctctgac 1570  
 Db 1567 TGGACTCCCATTTGGGAGAGGTGCTTTCTGCTGTGTATGAGCGGGAGCTGCTGAGC 1626  
 QY 1571 ggcattgtgcatgtgtgaaagggagacagatcatgycatcaacgcaagattgtgacag 1630  
 Db 1627 GGCATGCTGCAATTGTGAAGGGGAGAGATCATGCAATCAACGGCAAGATTGTGACAG 1686  
 QY 1631 actacacccctgctgagagcctgacgctgctcgtgcaagagcctggaatcagaagcgggagc 1690  
 Db 1687 ACACACCCCTGCTGAGAGCGCTGCCCTCTGCAAGAGGCTGGAATCAGGCGGAGCT 1746  
 QY 1691 ggaatgacactgtgtgttcgcgtctgcccccaagagagatgtagatgagctgaccttc 1750  
 Db 1747 GGATCGACCTTGTGTTGCGCTGTGCCCCCAAGAGGTATGACGATGAGCTGACTTCT 1806  
 QY 1751 tgcattgaagctcaaaagggaacccaattcaacgctttagaagaacagtgagctccggcccc 1810  
 Db 1807 T-CTGAAGTCCAAAGGGGAGAACCAATTCAC-AGTTAGGAAACAGTAGCTCCGGCCCC 1864  
 QY 1811 acctcgtgaaacaaagcctcgagccagccttggagagagcacaataacacacagat 1870  
 Db 1865 ACCTCGTGAACACAAACCTCGGTGATCAGCTTGGAGAGGCGCACACACACACAGAT 1924  
 QY 1871 ggcattccttggagactaatcatcaccacaggaatcacaactcccttggcccttgacc 1930  
 Db 1925 GGCATCTTGGAGACCTGAATCTATACCCAGAAATCTCAAACTCCCTTTGGCCCTGAAAC 1984

QY 1931 agggcagataaagaaacagctcgggccacttttttgaagcccaatgtgaggaagagag 1990  
 Db 1985 AGGGCCAGATAAAGGAACAGCTCGGGCCACTTCTTGAAGGCGCAACCTGAGGAAGAGAG 2044  
 QY 1991 cagccagccgcttggagaaagatcacaagatcagaactcactccttccttccttcgccc 2050  
 Db 2045 CAGCCACCATTTGGAGGAAGATCTCAAGATCCAGACTCTCATTTCTTCTTCTGCCCC 2104  
 QY 2051 agtgaattgtctctcccccagcttggggagctcctcttgaaccctaaataagaccaca 2110  
 Db 2105 AGTGAATTTGGTCTCTCCAGCTGTGGGGA-TCCTTCTCTGAACCTTAATGAACCCCA 2163  
 QY 2111 ctggagctctctctctccatccctcctcctcctcctcctcctcctcctcctcctcct 2170  
 Db 2164 CTGAGTCTCTCTCTCTCACTCCCTCTCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCT 2223  
 QY 2171 gtacatccaaacctactctcgagctcatataaataaacaagattatcttcagactta 2230  
 Db 2224 GTCACTCCAAACCTTACT-----GCACTCATGCAAAATAATTAATTAACCACTAA 2277  
 QY 2231 aaaaaa 2236  
 Db 2278 AAAAAA 2283  
 RESULT 5  
 AC005137/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 161190)  
 Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,  
 Buettner,J., Bumeister,R., Card,P., desaliboat,C., Dunn,J.,  
 English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,  
 Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,  
 Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,  
 Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.  
 HTGS Submission  
 Unpublished  
 2 (bases 1 to 161190)  
 Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,  
 Buettner,J., Bumeister,R., Card,P., desaliboat,C., Dunn,J.,  
 English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,  
 Grant,O., Hahner,L., Harris,J., Lewis,E., Loo,H., Loo,K.N.,  
 Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,  
 Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.  
 Direct Submission  
 Submitted (19-JUN-1998) Genome Science & Technology Center,  
 University of Texas Southwestern Medical Center, 5323 Harry Hines  
 Blvd., Dallas, TX 75235-8591, USA  
 3 (bases 1 to 161190)  
 Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,  
 Buettner,J., Butler,C., Card,P., desaliboat,F., Dunn,J.,  
 English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,  
 Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,  
 Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,  
 Schageman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.  
 Direct Submission  
 Submitted (30-SEP-1998) Genome Science & Technology Center,  
 University of Texas Southwestern Medical Center, 5323 Harry Hines  
 Blvd., Dallas, TX 75235-8591, USA  
 On Oct 1, 1998 this sequence version replaced gi:3241950.  
 Further information regarding the map of this region or  
 annotation of 6-106f23 can be found at  
 http://gscstc.somed.edu/chromosome2.htm.  
 CHROMOSOMAL LOCUS: This PAC clone comes from the Usher syndrome

region (USHC) mapped between STS markers D1S1310 and 11S1A14. MARKER CONFIRMATION: STS/ESTs sequence confirmed; D1S1032 and D1S4138. MAPPED CLONE OVERLAP: pD1239b22 and 6-109b6. IMPORTANT: This submission contains the entire insert of clone 6-106f23. 6-106f23 comes from a PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.40. In addition, this sequence has been finished such that 99.9% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

```
FEATURES
Source          Location/Qualifiers
                1..161190
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
repeat_region   574..955
                /rpt_family="MLT282"
repeat_region   1445..1635
                /rpt_family="MIR"
repeat_region   3125..3302
                /rpt_family="MIR"
repeat_region   11175..11329
                /rpt_family="MIR"
repeat_region   14552..14716
                /rpt_family="MIR"
repeat_region   complement(18589)..18627)
                /rpt_family="MIR"
repeat_region   23165..23232
                /rpt_family="MIR"
repeat_region   complement(25103)..25361)
                /rpt_family="Alu"
repeat_region   29622..29909
                /rpt_family="Alu"
repeat_region   complement(32772)..32989)
                /rpt_family="Alu"
repeat_region   33175..33284
                /rpt_family="MIR"
repeat_region   complement(34762)..34852)
                /rpt_family="MIR"
repeat_region   34864..34996
                /rpt_family="MIR"
repeat_region   complement(35719)..35843)
                /rpt_family="Alu"
repeat_region   complement(36294)..36467)
                /rpt_family="MIR"
repeat_region   36817..37084
                /rpt_family="Alu"
repeat_region   37953..38186
                /rpt_family="Alu"
repeat_region   39000..39177
                /rpt_family="MSR1"
repeat_region   complement(39898)..39968)
                /rpt_family="MIR"
repeat_region   complement(41579)..41780)
                /rpt_family="MIR"
repeat_region   complement(43317)..43597)
                /rpt_family="Alu"
repeat_region   45609..45899
                /rpt_family="Alu"
repeat_region   complement(46819)..46884)
                /rpt_family="MIR"
repeat_region   complement(48640)..48895)
                /rpt_family="MIR"
repeat_region   50385..50543
                /rpt_family="MIR"
repeat_region   complement(50842)..51126)
                /rpt_family="Alu"
repeat_region   complement(53140)..53180)
                /rpt_family="MIR"
repeat_region   complement(55758)..56060)
                /rpt_family="MIR"
```

```
repeat_region   /rpt_family="Alu"
                complement(59724)..60021)
repeat_region   /rpt_family="Alu"
                66550..66645
repeat_region   /rpt_family="MIR"
                73180..73260
repeat_region   /rpt_family="Alu"
                complement(74694)..74990)
repeat_region   /rpt_family="Alu"
                complement(76214)..77037)
repeat_region   /rpt_family="MER42"
                complement(77130)..77414)
repeat_region   /rpt_family="MER2"
                complement(77532)..77641)
repeat_region   /rpt_family="MER42"
                complement(78518)..78630)
repeat_region   /rpt_family="Alu"
                complement(78918)..79211)
repeat_region   /rpt_family="Alu"
                85547..85650
repeat_region   /rpt_family="MIR"
                88495..88613
repeat_region   /rpt_family="MIR"
                complement(93509)..93900)
repeat_region   /rpt_family="LI"
                complement(94148)..94417)
repeat_region   /rpt_family="Alu"
                complement(95536)..95822)
repeat_region   /rpt_family="Alu"
                complement(96013)..96567)
repeat_region   /rpt_family="LI"
                complement(106391)..106583)
repeat_region   /rpt_family="MLT1"
                complement(114951)..115156)
repeat_region   /rpt_family="MIR"
                118500..118580
repeat_region   /rpt_family="MIR"
                complement(126403)..126593)
repeat_region   /rpt_family="LI"
                126665..126942
repeat_region   /rpt_family="Alu"
                complement(127152)..127352)
repeat_region   /rpt_family="Alu"
                complement(128074)..128639)
repeat_region   /rpt_family="LI"
                complement(128923)..129156)
repeat_region   /rpt_family="Alu"
                130009..130278
repeat_region   /rpt_family="Alu"
                complement(134144)..139342)
repeat_region   /rpt_family="LI"
                complement(137205)..137469)
repeat_region   /rpt_family="Alu"
                complement(140667)..140762)
repeat_region   /rpt_family="MIR"
                complement(142407)..142521)
repeat_region   /rpt_family="MER42"
                147830..148006
repeat_region   /rpt_family="MIR"
                149821..150066
repeat_region   /rpt_family="Alu"
                complement(151137)..151327)
repeat_region   /rpt_family="MIR"
                160309..160771
repeat_region   /rpt_family="MLT1"
                37674 a 40113 c 40195 g 43208 t
```

BASE COUNT 37674 a 40113 c 40195 g 43208 t

Query Match 19.6%; Score 437.4; DB 11; Length 161190;  
Best Local Similarity 95.7%; Pred. No. 3.7e-90;  
Matches 472; Conservative 0; Mismatches 16; Indels 5; Gaps 2;

|    |      |  |      |
|----|------|--|------|
| Oy | 1740 | gcacacctctctgtcgaagtcacaaagggaaccaaattcaacgctttagaagaacactga   | 1799 |
| Db | 5253 | GCAGGACCTTTCTTGANGTCCAAAAGGGAAACCAAAATTTCAC-CGTTAGAAAACAGTA    | 5195 |
| Oy | 1800 | gtctcggcccaactcgtgtacacaaagccttcgagaccgctctagaagagccacatgac    | 1859 |
| Db | 5194 | GCTCGGGCCCACTCGTGAACACAAAGCCCTGGATCAGCTTGAGAGGCCACTAC          | 5135 |
| Oy | 1860 | acacacagatgtgcatacctctggagacctgaatctatacccaaggaaatctcaactccctt | 1919 |
| Db | 5134 | ACACACCGAGATGGCATCTCTGGGACCTGAATCTATCACCCAGGAATCTCAAACTCCCTT   | 5075 |
| Oy | 1920 | ggccctgaaccaggccagataaagaacagcttgggcacttllttagaagccaatgttg     | 1979 |
| Db | 5074 | GGCCCTGAACCAAGGGCCAGATAGGAACGCTCGGGCCACTCTTCTAAGGCCAACCTGG     | 5015 |
| Oy | 1980 | aggaagaagagcagccacgcgtttggagaagatctcaagaatccagacttaactctt      | 2039 |
| Db | 5014 | AGGAAGGGACAGCCACCCATTGGGGAAGATCTCAAGATCCAGATCTCTCAATTCCTT      | 4955 |
| Oy | 2040 | tcctctggcccgatgaatttgtctctccagacttggggagcttccttccttgaaaccta    | 2099 |
| Db | 4954 | TCTCTGGCCCAAGTAATTTGGTCTCTCTCCAGCTCTGGGGACATCCCTTCCTGAACCTTA   | 4895 |
| Oy | 2100 | ataagaaccacatgagctctctctctccatccctctcctctgcccctgtgcttaatlg     | 2159 |
| Db | 4894 | ATAAGACCCCACTGGAGTCTCTCTCTCCATCCCTCTCCTCTGCTCTTAATTTG          | 4835 |
| Oy | 2160 | ctgcacagatgtcaatcccaaccttactctcgagctcaatataataaacaagatttat     | 2219 |
| Db | 4834 | CTGCACGAGATTGTCACTCCAAACCTTACTCTGAGCTCATTAATAAATAAT---AGATTAT  | 4779 |
| Oy | 2220 | tttcacgcttaaa 2232   |      |
| Db | 4778 | TTTCCAGCTTATA 4766   |      |

|             |  |
|-------------|--|
| RESULT      | 6  |
| G55504/c    |  |
| LOCUS       |  |
| DEFINITION  | G55504 463 bp DNA  |
| DESCRIPTION | SHGC-100852 Human Homo sapiens STS genomic, sequence tagged site |
| REMARKS     |  |

|           |   |
|-----------|---|
| ACCESSION | G55504  |
| VERSION   | G55504.1  |
| KEYWORDS  | STs.  |
| SOURCE    | human.  |
| ORGANISM  | Homo sapiens  |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. |
| AUTHORS   | Myers, R.M.   |
| TITLE     | Human STS (1999)  |
| JOURNAL   | Unpublished (1999)  |
| COMMENT   |   |

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu  
Primer A: TCCAAAGTCAGGAAATGACATCT  
Primer B: TGTAAGAAGCAACGAGATCATGG  
STS size: 313  
PCR Profile:  
Initial Incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM

dNTPs: each 200 uM  
 AmpliTaq Gold Polymerase: 0.07 units/ul  
 Total Vol: 5 ul

Buffer:

|                     |        |
|---------------------|--------|
| MgCl <sub>2</sub> : | 2.5 mM |
| KCl:                | 50 mM  |
| Tris-HCl:           | 10 mM  |
| pH:                 | 8.3    |

BAC ends sequenced at TIGR from the RPC11 BAC library. Designed and developed at the Stanford Human Genome Center.

FEATURES

source

1. .463  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="11"  
 /clone\_id="Human"  
 80. .392  
 80. .102  
 primer\_bind  
 complement(370. .392)  
 96 a 116 c 128 g 123 t

BASE COUNT  
 ORIGIN

| Query Match | Similarity | 4.7%  | Score 105       | DB 13      | Length 463 |
|-------------|------------|---|-----------------|------------|------------|
| Best Local  | Similarity | 82.8%   | Pred. No. 6e-14 |            |            |
| Matches     | 120        | Conservative  | 0               | Mismatches | 25         |
|             |            |   |                 | Indels     | 0          |
|             |            |   |                 | Gaps       | 0          |
| QY          | 1576       | ggtggcattgtgaaaggagacagatcatgycgaatcaacgycgaagattgacagactac | 1635            |            |            |
| DB          | 400        | ggtgcattgtgaaaggagacagatcatgycgaatcaacgycgaagattgacagactac  | 341             |            |            |
| QY          | 1636       | acccttgctgaagctggaagctgcctgcgaagagccttgatcaagggcggaactggatc | 1695            |            |            |
| DB          | 340        | acccttgctgaagctggaagctgcctgcgaagagccttgatcaagggcggaactggatc | 281             |            |            |
| QY          | 1696       | gacctgtgattgtgcgtctgcgcccc                                  | 1720            |            |            |
| DB          | 280        | aggcccttcctccttctccctcc                                     | 256             |            |            |

|  |            |  |         |     |     |             |
|--|------------|--|---------|-----|-----|-------------|
|  | RESULT     | 7  |         |     |     |             |
|  | LOCUS      | 166494/c   |         |     |     |             |
|  | DEFINITION | I66494   | 7218 bp | DNA | PAT | 28-DEC-1997 |
|  | ACCESSION  | Sequence 14 from patent US 5670367.                            |         |     |     |             |
|  | VERSION    | I66494   |         |     |     |             |
|  | KEYWORDS   | I66494.1 GI: 2724471   |         |     |     |             |
|  | SOURCE     | .  |         |     |     |             |
|  | ORGANISM   | Unknown.   |         |     |     |             |
|  | REFERENCE  | Unclassified.  |         |     |     |             |
|  | AUTHORS    | 1 (bases 1 to 7218)  |         |     |     |             |
|  | TITLE      | Donner,F., Scheiflinger,F. and Falkner,F.Gunter.               |         |     |     |             |
|  | JOURNAL    | Recombinant Iowipox virus Patent: US 5670367-A 14 23-SEP-1997; |         |     |     |             |
|  | FEATURES   | Location/Qualifiers  |         |     |     |             |
|  | source     | 1..7218 /Organism="unknown"                                    |         |     |     |             |
|  | BASE COUNT | 1944 a 1491 c 1486 g 1929 t                                    |         |     |     | 368 others  |
|  | ORIGIN     |  |         |     |     |             |

```

Query Match          3.6%; Score 81.4; DB 5, length 7218;
Best Local Similarity 4.7%; Pred. No. 1,9e-08;
Matches 19; Conservative 246; Mismatches 142; Indels 0; Gaps 0

QY 1032 gcaagcgacagagcctctcatcgaagaagcgctgcgatggatccaagaatctcca 1091
      | | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1438 GTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1379

QY 1092 ggaagcagcagagatgagcgcgcaagaagaataattgccagaagcagcagaagga 1151

```

[illegible]

|             |   |                 |
|-------------|---|-----------------|
| RESULT      | 8   |                 |
| AC0007461.c |   |                 |
| LOCUS       | AC0007461   |                 |
| DEFINITION  | Homo sapiens chromosome 17, clone 84-E_24, complete sequence. | PRI 04-JUN-1999 |
| ACCESSION   | AC0007461   |                 |
| VERSION     | AC0007461.8   | GI:4996920      |
| KEYWORDS    | HTG.  |                 |
| SOURCE      | human.  |                 |

| ORGANISM  | Homo sapiens   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; |
|-----------|--|---|
| REFERENCE | 1 (bases 1 to 180385)  | Eutheria; Primates; Catarrhini; Homiidae; Homo.               |
| AUTHORS   | Biten,B., Linton,L., Nusbaum,C. and Lander,E.                      |   |
| TITLE     | Homo sapiens chromosome 17, clone 84_E_24                          |   |
| JOURNAL   | Unpublished  |   |
| REFERENCE | 2 (bases 1 to 180385)  |   |
| AUTHORS   | Biten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., |   |

| TITLE             | JOURNAL                 | REFERENCE  | AUTHORS  |
|-------------------|-------------------------|--|--|
| Direct Submission | Submitted (04-MAY-1999) | Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA | 3 (bases 1 to 180385)  |
|                   |                         |  | Birtner, B., Binton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., |
|                   |                         |  | Baker, J., Baldwin, J., Barna, N., Beckery, R., Benn, J., Brown, A.,       |
|                   |                         |  | Castell, A., Cerny, J., Colangelo, M., Collins, S., Collumore, A.,         |
|                   |                         |  | Cooke, P., Dardellano, K., Depayre, E., Devon, K., Dewar, K.,              |
|                   |                         |  | Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, M., Forrest, C.,           |
|                   |                         |  | Funke, R., Gage, D., Galagan, J., Gaidyna, S., Gilbert, R., Grant, G.,     |
|                   |                         |  | Hagos, B., Heaford, A., Horton, L., Howland, J., C. Jones, C., Kann, L.,   |
|                   |                         |  | Karatsas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,             |
|                   |                         |  | Marquis, N., McEwan, P., McGuck, A., McKernan, K., McLaughlin, J.,         |
|                   |                         |  | Meldrum, J., Molla, M., Morris, M., Morrow, J., Mychaleckyj, J.,           |
|                   |                         |  | Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,           |
|                   |                         |  | Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,    |
|                   |                         |  | Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,            |
|                   |                         |  | Testaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,      |
|                   |                         |  | Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.                     |

Direct Submission  
Submitted (04-MAY-1999) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 180385)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,  
Baker, J., Baldwin, J., Barna, N., Beckery, R., Benn, J., Brown, A.,  
Castle, A., Cerny, J., Colangelo, M., Collins, S., Collumore, A.,  
Cooke, P., DeBartellano, K., Deyre, E., Devon, K., Dewar, K.,  
Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,  
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,  
Hagos, B., Heathcote, A., Horton, L., Howland, J., C. Jones, C., Kann, L.,  
Karas, A., Lenoci, J., Liu, C., Locke, K., MacDonald, P.,  
Marquis, N., McKean, P., McQuirk, A., McKernan, K., McLaughlin, J.,  
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,  
Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,  
Peterson, K., Pollara, J., Riley, R., Roberts, D., Roy, A., Severy, P.,  
Strange-Thomann, N., Stojanovic, N., Stone, C., Sudramanlan, A.,  
Tessier, S., Tortorella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,  
Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

| TITLE         | COMMENT  |
|---------------|--|
| Journal       | Direct Submission  |
| Journal       | Submitted (04-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA   |
| Journal       | On Jun 4, 1999 this sequence version replaced gi:4966400.  |
| Journal       | All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)   |
| Journal       | <a href="http://ftp.genome.washington.edu/RW/RepeatMasker.html">http://ftp.genome.washington.edu/RW/RepeatMasker.html</a>  |
| FEATURES      | Only 180385 base pairs from the middle of this clone are being submitted. The remainder overlaps either accession number AC0051523 (WICR project L335) or accession number AC005243 (WICR project L343). |
| Source        | Location/Qualifiers  |
|               | 1.180385   |
|               | /organism="Homo sapiens"   |
|               | /db_xref="taxon:9606"  |
|               | /clone="84.E.24"   |
|               | /clone_lib="Alan Buckler -- per comm"  |
|               | /map="17"  |
|               | /chromosome="17"   |
| repeat_region | 446..753   |
| repeat_region | /rpt_family="AluSc"  |
| repeat_region | 2375..2419   |
| repeat_region | /rpt_family="(TTTG)n"  |
| repeat_region | complement(2985..3103)   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | complement(3146..3352)   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | complement(3531..3639)   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | complement(3796..4394)   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | 5531..5631   |
| repeat_region | /rpt_family="MER81"  |
| repeat_region | 5648..5674   |
| repeat_region | /rpt_family="(CAAAA)n"   |
| repeat_region | 5812..6020   |
| repeat_region | /rpt_family="L1MC4"  |
| repeat_region | 6180..6224   |
| repeat_region | /rpt_family="AT_rich"  |
| repeat_region | 6303..6611   |
| repeat_region | /rpt_family="AluSp"  |
| repeat_region | 7378..7618   |
| repeat_region | /rpt_family="MIR"  |
| repeat_region | 9593..9623   |
| repeat_region | /rpt_family="(GGGA)n"  |
| repeat_region | complement(10665..10923)   |
| repeat_region | /rpt_family="AluSq"  |
| repeat_region | 10958..11068   |
| repeat_region | /rpt_family="MIR"  |
| repeat_region | 11752..11914   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | complement(11915..12219)   |
| repeat_region | /rpt_family="AluX"   |
| repeat_region | 12220..12347   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | 13110..13182   |
| repeat_region | /rpt_family="GA-rich"  |
| repeat_region | 13526..13564   |
| repeat_region | /rpt_family="(TCC)n"   |
| repeat_region | 13575..13660   |
| repeat_region | /rpt_family="(TA)n"  |
| repeat_region | complement(14380..14658)   |
| repeat_region | /rpt_family="AluDo"  |
| repeat_region | 14742..14830   |
| repeat_region | /rpt_family="(TA)n"  |
| repeat_region | 14860..14967   |
| repeat_region | /rpt_family="(TA)n"  |
| repeat_region | complement(14977..15122)   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | complement(15292..15385)   |
| repeat_region | /rpt_family="L2"   |
| repeat_region | 16672..16781   |

|                              |  |  |                      |
|------------------------------|--|--|----------------------|
|                              | repeat_region  | 38083..38262   | /rpt_family="(TTC)n" |
|                              | repeat_region  | 38235..38319   | /rpt_family="(TTC)n" |
|                              | repeat_region  | 38320..38499   | /rpt_family="(TTC)n" |
| Query Match                  |  | 3.1%; Score 69.6; DB 40; Length 180385;                        |                      |
| Best Local Similarity        | 53.7%; Pred. Mismatches 124; Indels 0; Gaps 0                        |  |                      |
| Matches 144; Conservative 0; |  |  |                      |
| Oy                           | 978  | ggagctgttcattgaaaccgcggaggcgcttgcgaagggcgcgcacgtygactgcagcg    | 1037                 |
| Dd                           | 38451  | GGAGAAAGGAGGAAGAAGCAAGCAAGGAGGAAGGAAGGAGGAAGCAAGAAAGAAAGAA     | 38392                |
| Oy                           | 1038   | gcaagagcttcctcatgcacgaagcgcgtgcgatgtgaatccacaagaatccccaagga    | 1097                 |
| Dd                           | 38391  | GAAAGAGAAAGGAAGCAACAGCAAGCAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG     | 38332                |
| Oy                           | 1098   | gcaagagatgtgacgcgcgaagagaanaatlgtcccagaagcgcagcagaggaataatga   | 1157                 |
| Dd                           | 38331  | GAAAGAAAGGAAGGAAGCAAGCAAGCAAGCAGAGCAGAGCAGAGCAGAGCAGAGCAAG     | 38272                |
| Oy                           | 1158   | gagataccggaaaggagatygaaacagatttgtagagggaggaagaaatttaagaagcaatg | 1217                 |
| Dd                           | 38271  | GGAGAAAGGAGGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG    | 38212                |
| Oy                           | 1218   | gnaagaagactgtggcgtcaaaggaacg                                   | 1245                 |
| Dd                           | 38211  | GGAAGAGGAGGAGCGGAGCAAGCAAGCAAG                                 | 38184                |
| RESULT 9                     |  |  |                      |
| ACOL16063                    |  |  |                      |
| LOCUS                        | ACOL16063  | 151428 bp  | DNA                  |
| DEFINITION                   | Homo sapiens clone RP11-1706,  | WORKING DRAFT SEQUENCE,  | 12 unordered pieces. |
| ACCESSION                    | ACOL16063  |  |                      |
| VERSION                      | GI:6970444   |  |                      |
| KEYWORDS                     | HTG; HTGS_PHASE1; HTGS_DRAFT.  |  |                      |
| SOURCE                       | human.   |  |                      |
| ORGANISM                     | Homo sapiens   |  |                      |
| REFERENCE                    | Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;          |  |                      |
| AUTHORS                      | Euhneria; Primates; Carnivora; Homidae; Homo.                        |  |                      |
| TITLE                        | Britten,B., Linton,L., Nusbaum,C. and Lander,E.                      |  |                      |
| JOURNAL                      | Homo sapiens, clone RP11-1706  |  |                      |
| REFERENCE                    | Unpublished  |  |                      |
| AUTHORS                      | 2 (bases 1 to 151428)  |  |                      |
|                              | Britten,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., |  |                      |
|                              | Baldwin,J., Barne,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,   |  |                      |
|                              | Brown,A., Castle,A., Colangelo,M., Collins,S., Collamore,A.,         |  |                      |
|                              | Cooke,P., Dellallano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M.  |  |                      |
|                              | Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,             |  |                      |
|                              | Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,   |  |                      |
|                              | Howard,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,    |  |                      |
|                              | Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,            |  |                      |
|                              | Mckean,P., McGurt,A., McKernan,K., McLaughlin,J., Meldrum,J.,        |  |                      |
|                              | Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,        |  |                      |
|                              | Peterson,K., Pollara,V., Riley,K., Roy,A., Santos,R., Severy,P.,     |  |                      |
|                              | Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talmas,J.,         |  |                      |
|                              | Testeja,S., Tjirell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,      |  |                      |
|                              | Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.                             |  |                      |
| TITLE                        | Direct Submission  |  |                      |
| JOURNAL                      | Submitted (20-NOV-1999) Whitehead Institute/MIT Center for Genome    |  |                      |
| COMMENT                      | Research, 320 Charles Street, Cambridge, MA 02141, USA               |  |                      |
|                              | On Feb 14, 2000 this sequence version replaced gi:6721365.           |  |                      |
|                              | All repeats were identified using RepeatMasker:                      |  |                      |
|                              | Smit,A.F.A. & Green,P. (1996-1997)                                   |  |                      |
|                              | http://ftp.genome.washington.edu/RM/RepeatMasker.html                |  |                      |
|                              | -----Genome Center   |  |                      |
|                              | Center: Whitehead Institute/ MIT Center for Genome Research          |  |                      |
|                              | Center code: WIRB  |  |                      |

|               |                       |                           |
|---------------|-----------------------|---------------------------|
| repeat_region | /rpt_family="t2"      | 16802 . 16846             |
| repeat_region | /rpt_family="TGAA)n"  | 17225 . 17245             |
| repeat_region | /rpt_family="(A)n"    | 18411 . 18720             |
| repeat_region | /rpt_family="AluDb"   | complement(20715 . 20788) |
| repeat_region | /rpt_family="MLT1J2"  | complement(20822 . 20916) |
| repeat_region | /rpt_family="MLT1J2"  | 20956 . 21137             |
| repeat_region | /rpt_family="MIR"     | 21830 . 21872             |
| repeat_region | /rpt_family="AT-rich" | 21957 . 22257             |
| repeat_region | /rpt_family="AluSx"   | complement(22438 . 22538) |
| repeat_region | /rpt_family="MLT1J1"  | complement(23375 . 23667) |
| repeat_region | /rpt_family="AluY"    | 24553 . 24858             |
| repeat_region | /rpt_family="AluY"    | complement(25021 . 25234) |
| repeat_region | /rpt_family="MIR"     | 26957 . 27261             |
| repeat_region | /rpt_family="AluY"    | complement(27524 . 27669) |
| repeat_region | /rpt_family="MLT1B"   | complement(27870 . 27964) |
| repeat_region | /rpt_family="AluSp"   | complement(27965 . 28210) |
| repeat_region | /rpt_family="MLT1B"   | complement(28228 . 28522) |
| repeat_region | /rpt_family="AluSg"   | complement(28592 . 28700) |
| repeat_region | /rpt_family="MIR"     | complement(29533 . 29537) |
| repeat_region | /rpt_family="MIR"     | 29802 . 30095             |
| repeat_region | /rpt_family="AluSg"   | complement(30099 . 30219) |
| repeat_region | /rpt_family="MIR"     | 30227 . 30267             |
| repeat_region | /rpt_family="(TG)n"   | 30276 . 30321             |
| repeat_region | /rpt_family="(CA)n"   | complement(30322 . 30611) |
| repeat_region | /rpt_family="AluDo"   | complement(30956 . 31151) |
| repeat_region | /rpt_family="MER3"    | complement(31533 . 31633) |
| repeat_region | /rpt_family="t2"      | 31810 . 31858             |
| repeat_region | /rpt_family="t2"      | 32028 . 32107             |
| repeat_region | /rpt_family="L1PA4"   | complement(32919 . 33182) |
| repeat_region | /rpt_family="MIR"     | 34169 . 34314             |
| repeat_region | /rpt_family="CT-rich" | 34774 . 35007             |
| repeat_region | /rpt_family="MIR"     | 35589 . 35621             |
| repeat_region | /rpt_family="AT-rich" | 35763 . 36142             |
| repeat_region | /rpt_family="MER41A"  | 36143 . 36452             |
| repeat_region | /rpt_family="AluSg"   | 36453 . 36636             |
| repeat_region | /rpt_family="MER41A"  | 38011 . 38071             |
| repeat_region | /rpt_family="CT-rich" |                           |



12752 15275: contig of 2524 bp in length  
\* gap of unknown length  
\* 15276 20080: contig of 4805 bp in length  
\* gap of unknown length  
\* 20081 24494: contig of 4414 bp in length  
\* gap of unknown length  
\* 24495 31437: contig of 6943 bp in length  
\* gap of unknown length  
\* 31438 37373: contig of 5936 bp in length  
\* gap of unknown length  
\* 37374 44749: contig of 7376 bp in length  
\* gap of unknown length  
\* 44750 50716: contig of 5967 bp in length  
\* gap of unknown length  
\* 50717 56335: contig of 5619 bp in length  
\* gap of unknown length  
\* 56336 66076: contig of 9741 bp in length  
\* gap of unknown length  
\* 66077 76677: contig of 10601 bp in length  
\* gap of unknown length  
\* 76678 94859: contig of 18182 bp in length  
\* gap of unknown length  
\* 94860 118422: contig of 23563 bp in length  
\* gap of unknown length  
\* 118423 142659: contig of 24237 bp in length  
\* gap of unknown length  
\* 142660 166171: contig of 23512 bp in length.  
Location/Qualifiers  
1. 166171  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="26\_F\_02"  
/clone\_1lb="RPCI-11 Human Male BAC"  
BASE COUNT 49124 a 37104 c 36032 g 42814 t 1097 others  
ORIGIN

Query Match 3.1%; Score 68.8; DB 42; Length 166171;  
Best Local Similarity 52.9%; Pred. No. 1.8e-05;  
Matches 148; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

QY 976 cggagactgttcataagacccggagcggctgacgaagcggcgacgctgagctgcag 1035  
|||||  
Db 130472 CGGAGCAGCAGAGGGGCGAGGCGGAGCAGAGGGCGCGGCGGAGGAGGCGAG 130531  
|||||  
QY 1036 cggcagagacttcataagacccggagcggctgacgaagcggcgacgctgagctgcag 1095  
|||||  
Db 130532 GTGCGGAGCAGCAGAGGGGCGAGGCGGAGCAGAGGGCGCGGCGGAGGAGGAG 130591  
|||||  
QY 1096 cagcagagatgagcggcgaaggaagaagaattgccagaagcagcagaggaanaat 1155  
|||||  
Db 130592 CAGATGGGGAGCAGCAGAGGAGCAGATGCAGACAGAGAGCAGATGGGGAGCAGAGAG 130651  
|||||  
QY 1156 gagagataccggaagagatggaacagattgttagagaggaagaaagtttaagaagcaa 1215  
|||||  
Db 130652 GAGCAGATGCGGAGAGCGAGGAGCAGATGGGGAGCAGAGGAGCGAGTGCAGAAACAG 130711  
|||||  
QY 1216 tgggaagaagactcgggagctcaagaagcagctactcttgc 1255  
|||||  
Db 130712 GAGGAGCAGGTGCAGAAAGCAGAGAGAGCAGATGCGCAGAGC 130751  
|||||

RESULT 11  
KUSU52064  
LOCUS KSU52064 3489 bp DNA VRL 24-OCT-1996  
DEFINITION Kaposi's sarcoma-associated herpes-like virus ORF73 homolog gene,  
complete cds.  
ACCESSION U52064  
VERSION U52064.1 GI:1633571  
KEYWORDS Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.

ORGANISM Kaposi's sarcoma-associated herpesvirus  
Viruses: dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Kaposi's sarcoma-associated herpesvirus.  
REFERENCE  
AUTHORS Cesarman,E., Nador,R.G., Bai,F., Bohenzky,R.A., Russo,J.J.,  
Moore,P.S., Chang,Y. and Knowles,D.M.  
TITLE Kaposi's sarcoma-associated herpesvirus contains G protein-coupled  
receptor and cyclin D homologs which are expressed in Kaposi's  
sarcoma and malignant lymphoma  
J. Virol. 70 (11), 8218-8223 (1996)  
MEDLINE 97048116  
REFERENCE 2 (bases 1 to 3489)  
AUTHORS Bohenzky,R.A., Rusek,J., Edelman,I.S., Chang,Y., Moore,P.S.,  
Cesarman,E. and Knowles,D.M.  
TITLE Direct Submission  
Submitted (21-MAR-1996) Roy A. Bohenzky, Pathology, Columbia  
University College of Physicians and Surgeons, 630 W. 168th Street,  
New York, NY 10032, USA  
JOURNAL

FEATURES  
source location/Qualifiers  
1. 3489  
/organism="Kaposi's sarcoma-associated herpesvirus"  
/specific\_host="Homo sapiens"  
/db\_xref="taxon:37296"  
/clone="154"  
/cell\_line="BC-1, a body-cavity based lymphoma cell line"  
1. 3489  
/note="Herpesvirus saimiri ORF73 homolog"  
/codon\_start=1  
/protein\_id="AACS5944.1"  
/translation="MAPRGMRIRSGRSTGAPLRGSCRRNRSPERCDDGLDHLDR  
RKHVADSIDRECGPHLPPIGSPFTFSLPFTVSSPPLPVPATSPATPPLPPA  
LPPVTSSSPPIPSHPVSPGTTDHSPLPSPSSQNPPLSPSPGRDSSPT  
MRPSPQOTTPPHSPPTPPSPSPSPSLASTLRSLRSLSPSGSTLNPICQ  
SPVSPSPCDPFRSVYPPVATSPSPVSSSGDPPPPSPSPSPSPSPSPSP  
GDDTAMVLIAETAEASKNKEKCSNNQAQGEONGNETSKESQVKNDDNDEE  
OETDERDE  
DERDE  
EPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ  
PQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQEPQOQ  
REPOQREPOQREPOQREPOQREPOQREPOQREPOQREPOQREPOQREPOQ  
QOQ  
DEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDEQOQDE  
QOQ  
EOQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLE  
EOQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLEQOQLE  
EVEVEQOQ  
DQOQ  
DQOQ  
ARPQHIFTRFTRFTRFTRFTRFTRFTRFTRFTRFTRFTRFTRFTRFTRFTRF  
PHGPDQSPFTTCVIVYCONKDYSKRVQMARLAWASHPLAGNLQSSIVKFKPLPLT  
QPENGQPGDSPEMT"

BASE COUNT 1053 a 862 c 1137 g 437 t  
ORIGIN

Query Match 3.1%; Score 68.4; DB 16; Length 3489;  
Best Local Similarity 47.6%; Pred. No. 1.8e-05;  
Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 979 gagctgttcataagacccggagcggctgacgaagcggcgacgctgagctgcagcgg 1038  
|||||  
Db 2182 GAGCAGCAGCAGCAGATGACACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2241  
|||||  
QY 1039 cagagacttcataagacccggagcggctgacgaagcggcgacgctgagctgcagcgg 1098  
|||||  
Db 2242 CAGCAGCAGCAGATGACACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2301  
|||||  
QY 1099 caggaagatgagcggcgaaggaagaattgccagaagcagcagaggaanaatgag 1158  
|||||  
Db 2302 GAGTTAGAGAGCAGCAGAGCAGAGTACAGATCAGAGCAGCAGCAGCAGCAGCAGCAG 2361  
|||||  
QY 1159 agataccggaagagatggaacagattgttagagaggaagaaagtttaagaagcaatg 1218  
|||||

Db 2362 CAGGAGTTAGAGCAGCAGCAGAGTTAGAGCAGCAGCAGGATTAGAGAGCAG 2421  
||  
Qy 1219 gaagaagactggggtccaagaagacgactcttgccttaaaacctaactctgtggtga 1278  
||  
Db 2422 GAGCAGGAGTTAGAGCAGCAGCAGGAGTTAGAGCAGCAGCAGATTAGAGAG 2481  
||  
Qy 1279 caccagtagcccttcgcaagcaaaaglaatgacagggatggaacctgagctcgagccc 1338  
||  
Db 2482 CAGGAGCAGCAGGATTAGAGCAGCAGCAGGAGTTAGAGAGCAGCAGGTTGGAAGCA 2541  
||  
Qy 1339 gcaagatgacctggaatggaggaagcagggagcagggagcagagattccggaaatgg 1398  
||  
Db 2542 GAGCAGGAGTTGGAAGCAGCAGCAGCAGGAGCAGGAGCAGGAGTTAGAGAGGTGTGAG 2601  
||  
Qy 1399 ga 1400  
||  
Db 2602 GA 2603

RESULT 12  
AR065852/c  
LOCUS AR065852 32207 bp DNA PAT 29-SEP-1999  
DEFINITION Sequence 20 from patent US 5849564.  
ACCESSION AR065852  
VERSION AR065852.1 GI:5996068  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 32207)  
AUTHORS Chang, Y., Bohenzky, R.A., Russo, J.J., Edelman, I.S. and Moore, P.S.  
TITLE Poly peptides from Kaposi's sarcoma-associated herpesvirus, DNA  
encoding same and uses thereof  
JOURNAL Patent: US 5849564-A 20 15-DEC-1998;  
FEATURES  
source 1. 32207  
/organism="unknown"  
BASE COUNT 7229 a 9156 c 8713 g 7109 t  
ORIGIN

Query Match 3.1%; Score 68.4; DB 5; Length 32207;  
Best Local Similarity 47.6%; Pred. No. 2e-05;  
Matches 201; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

Qy 979 gagctgttcattacagaccggagggcgctgcagggcgcgagcgtgtagctgcagcg 1038  
||  
Db 19815 GAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGATGAGCAG 19756  
||  
Qy 1039 caggagcttcattacagaccggagggcgctgcagggcgcgagcgtgtagctgcagcg 1098  
||  
Db 19755 CAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGATGAGCAG 19696  
||  
Qy 1099 caggagctgcagggcgaggaaggaagaattcccgaggaagcagcagaggaagaatgag 1158  
||  
Db 19695 GAGTTAGAGAGCAGCAGCAGCAGGATTAGAGATCAGCAGCAGGATTAGAGAGCAGGAG 19636  
||  
Qy 1159 agatcccgaaagagatggaacagatttagagggaggaaggaagatttaagaagcaatg 1218  
||  
Db 19635 CAGGAGTTAGAGCAGCAGCAGCAGGATTAGAGAGCAGCAGCAGCAGGATTAGAGAGCAG 19576  
||  
Qy 1219 gaagaagactgggggtccaagaagcagctactcttccttaaaacctaactgctgaggtga 1278  
||  
Db 19575 GAGCAGGAGTTAGAGAGCAGCAGCAGGAGTTAGAGAGCAGCAGCAGCAGGATTAGAGAG 19516  
||  
Qy 1279 caccagtagcccttcgcaagcaaaaglaatgacagggatggaacctgagctcgagccc 1338  
||  
Db 19515 CAGGAGCAGCAGGATTAGAGAGCAGCAGCAGGAGTTAGAGAGCAGCAGGTTGGAAGCA 19456  
||  
Qy 1339 gcaagatgacctggaatggaggaagcagggagcagggagcagagattccggaaatgg 1398  
||  
Db 19455 GAGCAGGAGTTGGAAGCAGCAGCAGCAGGAGCAGGAGCAGGAGTTAGAGAGGTGTGAG 19396  
||

Qy 1399 ga 1400  
||  
Db 19395 GA 19394

RESULT 13  
KSU75698/c  
LOCUS KSU75698 137508 bp DNA VRL 03-MAY-1997  
DEFINITION Kaposi's sarcoma-associated herpesvirus long unique region, 80  
putative ORF's and kaposin gene, complete cds.  
ACCESSION U75698  
VERSION U75698.1 GI:2065526  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.  
JOURNAL Kaposi's sarcoma-associated herpesvirus  
MEDLINE Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
REFERENCE Gammaherpesvirinae; Rhadinovirus.  
AUTHORS 1 (bases 47193 to 47522; 133099 to 133729)  
Chang, Y., Cesarman, E., Pessin, M.S., Lee, F., Culpepper, J.,  
Knowles, D.M. and Moore, P.S.  
IDENTIFICATION of herpesvirus-like DNA sequences in AIDS-associated  
Kaposi's sarcoma  
J. Virol. 70 (1), 549-558 (1996)  
2 (bases 35021 to 55726)  
Moore, P.S., Gao, S.J., Dominguez, G., Cesarman, E., Lungu, O.,  
Knowles, D.M., Garber, R., Pellett, P.F., McGeoch, D.J. and Chang, Y.  
Primary characterization of a herpesvirus agent associated with  
Kaposi's sarcoma  
J. Virol. 70 (1), 549-558 (1996)  
3 (bases 28661 to 29741; 117919 to 118101)  
Zhong, W., Wang, H., Herndler, B. and Ganem, D.  
Restricted expression of Kaposi sarcoma-associated herpesvirus  
(human herpesvirus 8) genes in Kaposi sarcoma  
Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)  
96270595  
4 (bases 122794 to 123567; 123809 to 127297; 129372 to 130400;  
130551 to 134441)  
Cesarman, E., Nador, R.G., Bal, F., Bohenzky, R.A., Russo, J.J.,  
Moore, P.S., Chang, Y. and Knowles, D.M.  
Kaposi's sarcoma-associated herpesvirus contains G protein-coupled  
receptor and cyclin D homologs which are expressed in Kaposi's  
sarcoma and malignant lymphoma  
J. Virol. 70 (11), 8218-8223 (1996)  
97048116  
5 (bases 1 to 137508)  
Moore, P.S., Boshoff, C., Weiss, R.A. and Chang, Y.  
Molecular mimicry of human cytokine and cytokine response pathway  
genes by KSHV  
Science 274 (5293), 1739-1744 (1996)  
97094384  
6 (bases 1 to 137508)  
Russo, J.J., Bohenzky, R.A., Chien, M.-C., Chen, J., Yan, M.,  
Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y. and  
Moore, P.S.  
Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus  
(HHV8)  
Proc. Natl. Acad. Sci. U.S.A. 93 (25), 14862-14867 (1996)  
97121480  
7 (bases 1 to 137508)  
Russo, J.J., Bohenzky, R.A., Chien, M.-C., Chen, J., Yan, M.,  
Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y. and  
Moore, P.S.  
Direct Submission  
Submitted (17-Oct-1996) Dept of Pathology/Genome Center/Division of  
Epidemiology, Columbia University, 650 West 168th Street, New York,  
NY 10032, USA  
8 (bases 1 to 137508)  
Russo, J.J., Bohenzky, R.A., Chien, M.-C., Chen, J., Yan, M.,  
Maddalena, D., Parry, J.P., Peruzzi, D., Edelman, I.S., Chang, Y. and  
Moore, P.S.  
Direct Submission

REFERENCE  
AUTHORS  
TITLE





|            |  |  |        |
|------------|--|--|--------|
| Oy         | 1099   | caggagatgtagcgccaaagaagaagaatctgcccaagaacgcagcagaggaaaatgag      | 1158   |
|            |  |  |        |
| Db         | 124996   | GAGTTTAGAGACACAGGACGAGAGATTAGAGATCAGAGACACAGGACTTAAAGAGCGAGGAG   | 124997 |
|            |  |  |        |
| Oy         | 1159   | agataccggaaggaatatggaacaagattgtagaaggagaaagaagttaagaacatgag      | 1218   |
|            |  |  |        |
| Db         | 124936   | CAGGAGTTTAGAGACACAGGACGAGGAGTTAGAGAGACAGAGCAGAGCAGAGTGAGGAGCAG   | 124877 |
|            |  |  |        |
| Oy         | 1219   | gaagaagactggggctccaagaagaacagctaccttctgcccataaacatcacctctgagcta  | 1278   |
|            |  |  |        |
| Db         | 124876   | GAGCAGGAGATTAGAGACGACGAGCAGGAGTTTAGAGAGACACAGCAGAGCAGGATTAGGAGAG | 124817 |
|            |  |  |        |
| Oy         | 1279   | caccoccatcccccttcgcacagcacaagtatgtatcagggaggtggaacctgagctcgagccc | 1338   |
|            |  |  |        |
| Db         | 124816   | CAGGAGCAGAGACTTAGAGAGACGACGAGCGGAGTTTAGAGAGACACAGAGGTGGAGAAGCAA  | 124757 |
|            |  |  |        |
| Oy         | 1339   | gcagaatgacctgtagatgatgagccagggagagcagggagagcagattcttcggaatatgag  | 1398   |
|            |  |  |        |
| Db         | 124756   | GAGCAGGAGAGCTGGGAAGAGCAAGACGACGAGCAGGAAGACACAGAACTTAAGAGAGGTGGAG | 124697 |
|            |  |  |        |
| Oy         | 1399   | ga 1400  |        |
|            |  |  |        |
| Db         | 124696   | GA 124695  |        |
|            |  |  |        |
| RESULT     | 14   |  |        |
| LOCUS      | U93872/c   |  |        |
| DEFINITION | U93872 133661 bp DNA VRL 09-JUL-1997   |  |        |
| KEYWORDS   | Kaposi's sarcoma-associated herpesvirus glycoprotein M, DNA replication protein, glycoprotein, DNA replication protein, FLICE inhibitory protein and v-cyclin genes, complete cds, and tegument protein gene, partial cds. |  |        |
| ACCSSION   | U93872   |  |        |
| SOURCE     | U93872.1 GI:2246466  |  |        |
| ORGANISM   | .  |  |        |
| REFERENCE  | Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8. Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus. 1 (bases 29032 to 30108; 117733 to 118431)                           |  |        |
| AUTHORS    | Zhong, W., Wang, H., Herndler, B. and Ganem, D.  |  |        |
| TITLE      | Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)   |  |        |
| JOURNAL    | Proc. Natl. Acad. Sci. U.S.A. 93 (13), 6641-6646 (1996)  |  |        |
| MEDLINE    | 96270595   |  |        |
| REFERENCE  | 2 (bases 17242 to 17856)   |  |        |
| AUTHORS    | Neipel, F., Albrecht, J.C., Ensser, A., Huang, Y.Q., Li, J.J., Friedmann-Kien, A.E. and Fleckenstein, B.   |  |        |
| TITLE      | Human herpesvirus 8 encodes a homolog of Interleukin-6 J. Virol. 71 (1), 839-842 (1997)  |  |        |
| JOURNAL    | J. Virol. 71 (1), 839-842 (1997)   |  |        |
| MEDLINE    | 97138401   |  |        |
| REFERENCE  | 3 (bases 123309 to 124082)   |  |        |
| AUTHORS    | Li, M., Lee, H., Yoon, D.W., Albrecht, J.C., Fleckenstein, B., Neipel, F. and Jung, J.U.   |  |        |
| TITLE      | Kaposi's sarcoma-associated herpesvirus encodes a functional cyclin J. Virol. 71 (3), 1984-1991 (1997)   |  |        |
| JOURNAL    | J. Virol. 71 (3), 1984-1991 (1997)   |  |        |
| MEDLINE    | 97184528   |  |        |
| REFERENCE  | 4 (bases 122660 to 123226)   |  |        |
| AUTHORS    | Thome, M., Schneider, P., Hofmann, K., Fickenscher, H., Mehl, E., Neipel, F., Matzmann, C., Burns, K., Bodmer, J.L., Schrotter, M., Seifeld, C., Kramer, P.H., Peter, M.E. and Tschopp, J.                                 |  |        |
| TITLE      | Viral FLICE-inhibitory proteins (FLIps) prevent apoptosis induced by death receptors Nature 386 (6624), 517-521 (1997)   |  |        |
| JOURNAL    | Nature 386 (6624), 517-521 (1997)  |  |        |
| MEDLINE    | 97242415   |  |        |
| REFERENCE  | 5 (bases 1 to 133661)  |  |        |
| AUTHORS    | Neipel, F., Albrecht, J.C. and Fleckenstein, B.  |  |        |
| TITLE      | Cell-homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? J. Virol. 71 (6), 4187-4192 (1997)   |  |        |
| JOURNAL    | J. Virol. 71 (6), 4187-4192 (1997)   |  |        |
| MEDLINE    | 97296220   |  |        |

[illegible]

```
/codon_start=1
/protein_id="AAB62620.1"
/db_xref="GI:2246495"
/translation="MAKELAAVYADVSAIAMDICLSIADPATIDTKSLATITGKROS
LHGTLLPLLRONAHCCSGLSLEHLEHLEWADALATLGVCSRKLSPEHFSILHID
TCNKRVSANFNFGMMALEKLSLINDVEIFPKRLSVYCGISGSALEGELEVR
VGLRGISVPGDGLQVSNPLCEIOEVLCTNOGSILOAMPDAGICHICPAGC
PVRGLFENELKQGLQTPNSIPTTPGQSVRRQDELROSSIAMVGDHIFGEYTRSVL
EISNLIYWSGSHDTCDDGDRDCSHLASLTHEDAKRRVDIAGLGEGRYKHFED
CFRPSLETLFCGGLFSVEDTIESLQKDSAFYQOVNTTALQKONEERYRLKIL
AAGQLNMGKSTSCSPSEARQOLVGKKPEVYLDARQDELQYARVAGFEKLSDC
IRHOGHILSOTGLRLMGSVIYNEASALOHNFHRAOFISLPMQDLDVDCPEAGMK
YIKNSLICORLGEHVEHILILEFYKLITGSLSRHLFSPVPPVATACCEAEMKPH
OKMVSSEMTIPSEIPKDMIEPNNOFTSPENDINILQKRAEYIKELIVSLVLYNT
WERELKILTPQSPGSEEPKPAGLTTGLTLPETSAPLVLDKKYGLFKDLIALLY
HLQLSNHNDSQ"
8680..11217
/note="ORF 08, glycoprotein B (gB) homolog"
/codon_start=1
/protein_id="AAB62592.1"
/db_xref="GI:2246467"
/translation="MTPRSRLATLGVILLVCFCAGAHNSGDTFQTSPPRPGSSS
KAPTKGEASGSPVDYQFQVCSASITGELFRMLQCPDTPKDYHQEGILLYK
KNIVPHIFKRYRRK IATSVYRGLESATIKKELPREVPLVIESHMDSTQCSS
MKVNVGVENTFTDRDQVNTVTFLOPEGLTDNIQRFVQSQPVYARBPMPGIVRYRT
TVNCEIVDMARSAEYNYFVSLGDTVEYSPCYNCESSSTTPSKNGISVQVVLNH
TVVYSIDRGTSPQNRHIFETGATVLSMASSEKTTAVCSLALMKTFPRISIQTHDS
FHFVANEITATFTAPLTPVANFTDYTSLTSDINTTILNMSKALASTHVNGVQFH
TGGELVLMQPMASINLTHAQDGNPTSPSPASBPMTSASRRRSASTAAAGG
GSTDNLSYTOLOQAFYDKLRDGINOVLEESLRACRROVNLNMLWELSKINPTSVTA
YGRPVSAKFEVDAISVTECINVDOSSVNTRKSLRNSNGSCARPLNPKFNSNL
FPGQLARNELITLNNVETCKDCEHYFTTRNETLYKYDIALRTINTDISTLMT
IALNLSFIQNIIDKALIELYSSAKRLASSFDELEMERENYTHRIAGRELDNTI
DMKKEFRVRLDELSEIVADLGIGKTVNVAASVVTLCAGSLTGFINIKHPLGCMII
IVAILIILFMSLRNTTIAQAEVKMIIPVDVRRAPSGCAPREETKNTLLGMHQL
QSEERKADLKRKTSVQFRTANGRLRGLRGYPLNGSIDISPETG"
11344..14382
/note="ORF 09, DNA polymerase homolog"
/codon_start=1
/protein_id="AAB62593.1"
/db_xref="GI:2246468"
/translation="MDPFNPFDTPTRGPNNTVROPSPQSPVPSFETRYCLIPAC
QTPGRGVAAVDTTPPTPYQGRKREYVAGETGSIKTRRGARNAHMLFHHYD
IYETTTADRCEDVPSPFOTDILPSGVALKLRITLDGASCVNFRQRCYFTYLAQ
GVNLTHTVLOALQAGFARASCGFSTBYRKKILRAYDQVAKYKTLSSPMRLS
DLRTTCGEFEFSNVDAIRFVLDHGFSTFGWEGSNPAPRTQARSDMTLEFDCSWE
DKFIERTEMEMPTTILSPDIECMGKGPENATODDMIIQISCVLHYVNDKPYRM
LIGTGCDPLPGVEFEFPESEYDMLAALISMLRDVAYEFTTGYNINAFDLPYIARAT
QYDPRLODPTKIKTGSVEYVHOPRGSGDGNMRBOSKYGICYPIIDMYQCREKL
SUSDVRLDVAKQCLGRQKODISYKDIPLFKSGPGRNAVGVYCIDSVLMDLIR
FQTHVSEISIAKLAKIPTRRVLTLDGQIINFVLSGLIAAATIEGILVPKDDAVSGVG
AVYISPSGFEYDDPLVVDVPSLPSIIQAHNLCYTLLEPGLSLHPLHSPDEYEF
VLSGQVHFVYKHKHRESLAKLITVWLAKREIKRLTASCTDALTILDKLOAIKY
TCNAVYGFVAGVAGSILPCINIAETVTLQGMKREARQPEALISPRILGRLPIDY
SPDARKVITGDTSLFICMGFMNDSVSPAEELASITNTITFRPIKIEAKITKC
LHLTKRIVGVLSDDVKMKGVDLIRKTRCRVOKRSSVDLILREBSVAANKIL
SGQATVWYREGIPEGEFVYIIQVLSNHRCLCRSPVDKLTPTTLESRLADYKQN
LPHLYQKIOARQOEBELPOIHRIPIYVADPAPSLSEIALEHEVYKQHLRAVADLY
PDKLVAHANIIICLFQNNMTSATVAMLYNFDLPVFPFP"
14500..15756
/note="ORF 10, herpesvirus saimiri ORF 10 homolog,
conserved in other gamma-herpesviruses"
/codon_start=1
/protein_id="AAB62594.1"
/db_xref="GI:2246469"
/translation="MOTFAFLLIGDWEITVNSCFTCCSLTGGPLRSSGDTYRLAP
FSIDRLIRHAIIRGLVNTIEDLTHSGCAVAVADANATGNNARITAPRTYNNSEPI
GIWVRGPPQTRKEAIKFCIEFVSPUPRKMITYVRRGGLPGAEAEPELHSAEVL
PSRETLVQGLRSTSPRTYGFHSBPISPLSFSTIESGCVNVEDPDLKAEVLT
FTQGERLCKVYVNTNSTACKARVAFVPRPSROLWAGASPLITPLARFA
VVDCEKTIPOETTLRIOLFEQGANAGDAFVILGARETKFSVPAVILPGH
EHLIVNPOHPLTIORDITVGVAMACIYHPGCAASQAPYSFDCKESSHVGLFQIK
RPGGYCTPCCHVAIRADRHEPMOS"
```



/note="similar to OX-2 membrane glycoprotein"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /protein\_id="BAD46502.1"  
 /db\_xref="GI:5669895"  
 /translation="MHFEFDCPPRRVEGVYISFYLLGAPGRATIKTEEGSALON  
 LPPTVAVGTSGSVVAPPTTPSPASRSKMSISFLSPAVATILALCAVGA  
 ROGMRGSAALATCAITPADADIVSYWOKROGPVNAVATYSHSYGVOTQYRHN  
 ITCPGLAMSTVITIHNLADDECCYLCITENSPGROGSCACLEVTSPPGHQNVNTE  
 DADYVTCATATGPPPNVYTAALAMNNNASSTOEQPTSDGLTVAMRYLRPGDVTTS  
 ASPTSDMHESIISTPASTIOGPAHLHPAAGTILGAVITLVGLGIFALHCKRKG  
 GSPSTDDMDPLSTQ"  
 14111..15262  
 /gene="ORF\_74"  
 14111..15139  
 /gene="ORF\_74"  
 /note="G protein coupled receptor; similar to IL-8  
 receptor; transforming protein"  
 /codon\_start=1  
 /evidence=experimental  
 /product="v-GPCR"  
 /protein\_id="BAD46503.1"  
 /db\_xref="GI:5669896"  
 /translation="MAAEFLTFLDDESNMETLMSGIDVSGNFSLEYSVCEMTY  
 VPMYVNGVLSLFLINVLGNGLVYIFEKHSRAAGADILLGLTINSKCSISLA  
 EYVMTLFPRIISTGICREIIEFYVYIDITSVCSLVRYLLVAVTSRSPKRSKL

|                       |                 |                    |           |               |
|-----------------------|-----------------|--------------------|-----------|---------------|
| Query Match           | 3.08;           | Score 67;          | DB 16;    | Length 28559; |
| Best Local Similarity | 47.38;          | Pred. No. 4.2e-05; |           |               |
| Matches 202;          | Conservative 0; | Mismatches 225;    | Indels 0; | Gaps 0;       |

|    |      |  |      |
|----|------|--|------|
| OY | 976  | cggaagactgttccatgacacacccgagcgagtcgtgcagagcgccgacgctgagctgag     | 1035 |
| Db | 9847 | CAGGAGTTAGAGGAGCGACGAGGACGAGAGTTTAGAGGACGACGACGAGAGTTAGAGAGCAG   | 9788 |
| OY | 1036 | cggcaggaagcttctcaltgcagaaacgcygcttgcgtatgagttccaaacaatctctccagag | 1095 |
| Db | 9787 | GAGCAGGGATTAGAGGAGCGAGGACAGAGATTTAGAGACGACGACGAGATTAGAGAG        | 9728 |
| OY | 1096 | cagcagaagatctgagcgcgcaaaagaaagaattgtcccgaaagcgccagggaaat         | 1155 |
| Db | 9777 | CAGAGCGCGGATTAGAGGAGCGAGGACGAGGATTAGAGGACGACGACGAGAGTTAGG        | 9668 |
| OY | 1156 | gagagataccggaagagatctgacacagatctgtagaaggaagaaatlttaaaagcaa       | 1215 |
| Db | 9667 | GAGCAGGAGCGAGGATTAGAGGAGCAGGAGCAGAGGATTAGAGACGACGACGAGAGTTA      | 9608 |
| OY | 1216 | tgggaaagaagacttggggtctcaaaagaaacgactctcttgcataaacatcatctgtag     | 1275 |
| Db | 9607 | GAGGAGGACGAGGACGAGGATTAGAGGAGCGAGGACGAGGACGAGGATTAGAGAGGTGGA     | 9548 |
| OY | 1276 | gtacaccacgaataccctctgcagaagcaaatgtatgtacaggggttggaacctgagctgag   | 1335 |
| Db | 9547 | GAGCAGAGGACGAGGACGAGGAGGACGAGGATTTAGAGAGAGGTGAGGACCAAGACGAGAG    | 9488 |
| OY | 1336 | ccgcagaatgacactgtgagtagaagcaccgagggagccagggagacagagatttccggaatat | 1395 |
| Db | 9487 | CAGGAGGAGCGAGGAGGAGGAGATTAGAGAGAGGTGGAAGACGAGGAAGACGAGAGTTA      | 9428 |
| OY | 1396 | gaggaag 1402   |      |
| Db | 9427 | GAGGAGG 9421   |      |

Search completed: May 20, 2000, 01:07:30  
Job time: 14654 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2000, 21:03:16 ; Search time 6491.04 Seconds  
(without alignments)  
-232.593 Million cell updates/sec

Title: US-09-502-945-1  
Perfect score: 1552  
Sequence: 1 cttctgagatgcacgcgagaa.....aaatgacttttaagaaga 1552

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_om:\*  
5: gb\_om:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl1:\*  
10: gb\_pl2:\*  
11: gb\_pl3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_jn:\*  
21: em\_om:\*  
22: em\_om:\*  
23: em\_om:\*  
24: em\_om:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_pl:\*  
28: em\_ro:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID         | Description         |
|------------|-------|-------------|--------|---------------|---------------------|
| 1          | 1552  | 100.0       | 1552   | 40 AF039690   | AF039690 Homo sapi  |
| 2          | 504   | 32.5        | 1325   | 51 AF161348   | AF161348 Homo sapi  |
| 3          | 326.4 | 21.0        | 1198   | 11 HS079271   | U79271 Human clone  |
| 4          | 257.2 | 16.6        | 576    | 13 G27428     | G27428 human SFS S  |
| 5          | 90.6  | 5.8         | 7218   | 5 I66494      | I66494 Sequence 14  |
| 6          | 67.2  | 4.3         | 1686   | 24 E08995     | E08995 DNA encodin  |
| 7          | 65.8  | 4.2         | 180385 | 40 AC007461   | AC007461 Homo sapi  |
| 8          | 62.4  | 4.0         | 39103  | 35 CEI03368   | US3154 Caenorhabdi  |
| 9          | 62.4  | 4.0         | 193188 | 41 AC006884   | AC006884 Caenorhab  |
| 10         | 59    | 3.8         | 17522  | 12 M05MHC4H2S | M17440 Mus musculu  |
| 11         | 55.6  | 3.6         | 2554   | 7 SCYK1202W   | Z28201 S.cerevisia  |
| 12         | 55.6  | 3.6         | 3997   | 5 E14049      | E14049 gDNA which   |
| 13         | 55.4  | 3.6         | 5420   | 34 AF056936   | AF056936 Plasmodi   |
| 14         | 54    | 3.5         | 4000   | 7 D83006      | D83006 Saccharomyc  |
| 15         | 54    | 3.5         | 100000 | 9 AP000067    | AP000067 Homo sapi  |
| 16         | 53.2  | 3.4         | 37797  | 34 CET21B4    | Z81124 Caenorhabdi  |
| 17         | 53.2  | 3.4         | 69134  | 32 CEY57A10_4 | Continuation (5 of  |
| 18         | 53.2  | 3.4         | 110000 | 32 CEY38E10_1 | Continuation (2 of  |
| 19         | 52.8  | 3.4         | 110000 | 32 PFMA14P1_0 | AL034557 Plasmodiu  |
| 20         | 52.8  | 3.4         | 148278 | 40 AC006101   | AC006101 citb_338_  |
| 21         | 52.6  | 3.4         | 175630 | 56 AC023059   | AC023059 Mus muscu  |
| 22         | 52.4  | 3.4         | 20776  | 34 CEK03D10   | Z81561 Caenorhabdi  |
| 23         | 52.4  | 3.4         | 110000 | 32 CEY63D3_0  | Continuation (2 of  |
| 24         | 52.4  | 3.4         | 110000 | 32 CEY63D3_1  | AC006832 Caenorhab  |
| 25         | 52.2  | 3.4         | 22502  | 41 AC006832   | Continuation (2 of  |
| 26         | 51.8  | 3.3         | 120000 | 41 AC007287   | AC007287 Homo sapi  |
| 27         | 51.8  | 3.3         | 151428 | 55 AC016063   | AC016063 Homo sapi  |
| 28         | 51.6  | 3.3         | 4238   | 12 RNCNG41    | AJ000496 Rattus no  |
| 29         | 51.6  | 3.3         | 315277 | 43 AC018634   | AC018634 Homo sapi  |
| 30         | 51.4  | 3.3         | 149886 | 12 MMHC438N12 | AF049850 Mus muscu  |
| 31         | 50.8  | 3.3         | 879    | 13 CNS01JRG   | AL147405 Anopheles  |
| 32         | 50.4  | 3.2         | 720    | 9 AB007820    | AB007820 Homo sapi  |
| 33         | 50.4  | 3.2         | 1410   | 9 AB009843    | AB009843 Homo sapi  |
| 34         | 50.2  | 3.2         | 155364 | 52 AC022248   | AC022248 Homo sapi  |
| 35         | 50    | 3.2         | 227070 | 56 AC023888   | AC023888 Homo sapi  |
| 36         | 49.6  | 3.2         | 1512   | 9 HSKERELP    | X62571 H.sapiens m  |
| 37         | 49.4  | 3.2         | 3987   | 43 AC017190   | AC017190 Drosophila |
| 38         | 49.2  | 3.2         | 3371   | 4 GGCLAUSA    | X67778 G.gallus mR  |
| 39         | 49.2  | 3.2         | 256172 | 41 AC005139   | AC005139 Plasmodiu  |
| 40         | 49.2  | 3.2         | 308399 | 44 AC005140   | AC005140 Plasmodiu  |
| 41         | 49    | 3.2         | 116515 | 43 AC013527   | AC013527 Homo sapi  |
| 42         | 49    | 3.2         | 173690 | 43 AC011307   | AC011307 Homo sapi  |
| 43         | 49    | 3.2         | 178975 | 41 AC011302   | AC011302 Homo sapi  |
| 44         | 48    | 3.2         | 265640 | 57 AC008046   | AC008046 Homo sapi  |
| 45         | 48.8  | 3.1         | 64411  | 32 CEY63D3_3  | Continuation (4 of  |

ALIGNMENTS

RESULT 1  
AF039690 1552 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-8 (NY-CO-8) mRNA, partial cds.  
DEFINITION AF039690  
ACCESSION AF039690.1 GI:3170179  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
1 (bases 1 to 1552)  
Gordan,J.D., Yureci,O., Sahin,U., Pfrendrich,M., and Old,L.J.  
TITLE Characterization of human colon cancer antigens recognized by  
antibodies  
JOURNAL Int.J.Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 1552)  
AUTHORS Scanlan,M.J.  
TITLE Direct Submission  
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1. 1552  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colorectal carcinoma"  
<1..1552  
/gene="NY-CO-8"  
<1..1245  
/gene="NY-CO-8"  
/codon\_start=1  
/product="antigen NY-CO-8"  
/protein\_id="AAC18039.1"  
/translation="LIDASEKLTLYEKECELESOLKFLRNDLAEYORFCEDLKROL  
/KKEFLIANTCNVGLCLCKOHEAVLSOTHTVHMOTIEFLVERDLMALVSY  
RSIADTPOORBAAYROVKVOIOISEANPEKTKALIOCDOLKELERELKELIA  
SOOEKRAITEKDMKKEITKEREKMSKMLISNTQLOAEVYKKEISATNOLEE  
IOSOLASREMDYKVCGEKRYOLNKTNMEKDEKEREKREKRAKTNDLEKDOELK  
RIELDESKHLEQEOKALARECEKRLTELLEGESEHOLHTRSEIAOLSQEKRYTD  
KIGKLORNNELEEDOCVHGVRHETMKORLROLDKHSQATAOOLVOLSKONOLLIER  
OSLSEVDRLRQLPSMPOSDC"  
BASE COUNT 525 a 291 c 401 g 335 t  
ORIGIN  
Query Match 100.0%; Score 1552; DB 40; Length 1552;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1552; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ctctcgatgcatcgagaaagctaaacttacttatagtgagaaagtgtgaattgagaa 60  
DB 1 CTTCTGGATGCATCCGAGAGCTAAACTTACTTATGAGAAAAGTGTGAATTGAGAA 60  
QY 61 tcccaattgaagtttttggagaacgtagctgaatatagaagaactgttgaagctt 120  
DB 61 TCCCAATTGAAGTTTGGAGGAAGACTTGTGAATATACAGAACTTGTGAAGATCTT 120  
QY 121 aaagagaactaagaactaagaattctctctgctgctataactgttaaccgttgtgt 180  
DB 121 AAAGAGCACTAAAGCTTAAGAATTCTTCTGCTGCTTAATCTGTAAACCGTGTGT 180  
QY 181 ggtcttgtttgaatgtgtcagcatgaagcgtgttcttcccaaacctactaatgt 240  
DB 181 GGCTTTGTTGAATGTGCTCAGCATGAGAGCTGTCTTCCCAAAACCATTAATGTT 240  
QY 241 catatgcagacatctgaaagactggttaaaagaagaagtactgtatgtctgactagt 300  
DB 241 CATATGCAGACATCTGAAAGACTGTGTTAAAGAAAGAGATGATTGATGTGCACTAGTT 300

QY 301 tccgtaagagcagcttgcgagatcacgcagcaagaagaagtgctta tgaaccagtg 360  
DB 301 TCCGTAAGAGAGCAGCTTGGCAGATACGCACAGAAAGAGCAAGAGTCTTATGAACAGGTG 360  
QY 361 aaacaagtttgcacaatcttgaaagaaagcaatttgaaaaaaaccaagccttaaccag 420  
DB 361 AAACAAGTTTGGCAAAATATCTGAGAGAGCCAAATTTTAAAAAACCAAGCGCTTTAATCTCAG 420  
QY 421 tgtgaccagtttgaagagagccttgagagagcagcgagcagacttgaagaagacttgc 480  
DB 421 TGTGACCAGTTTGAAGAGAGAGCTTGAGAGAGCGGAGCGACTTGAAGAAAGAACTTGCA 480  
QY 481 tctcagaagaagaagaagccattgagaagaacatga tgaaaaaagaaataacgaagaa 540  
DB 481 TCTCAGCAAGAGAAAAGGCCATTGAGAAAAGACATGATAAAAAAGAAATACGAAGAA 540  
QY 541 agggagatcacatggtatcaaaagtgttacttcttctagaatatctcccaactggaagcc 600  
DB 541 AGGGAGATCACATGGGATCAAAAGATGTTGATCTTCTCAGAAATATTTGCCCAACTGAGAGCC 600  
QY 601 cagtgtaaaaaagtttaaaaaagaagatttcagctataactgagaggaattcaa 660  
DB 601 CAGTGTAAGAAAGTTTACAAAGAAAGATTTCAGCTATTATCAACTGAGAGAAATTCAA 660  
QY 661 agccagctggtctctcggaatatgatatcaaaagtgctgtgagaagaatgctatcag 720  
DB 661 AGCCAGCTGGCTTCTCGGAAATGATGATCACAAGAGTGTGTGAGAGAAATGCCGTATCAG 720  
QY 721 ctgaataaacaacacatgtgagaagagatgaagcagaagaagaagcagagatcagaaga 780  
DB 721 CTGAATTAACCAACATGAGAGAGATGAGAGCAAGAAAGAGCACAGAGAGTTCAGAGCA 780  
QY 781 aaacttaacagagatcttgaataataagatcaggaataagaagaattgagaatagaactg 840  
DB 781 AAACCTAACAGGAGTCTTGAATTAAGATCAGAGAAATTAAGAAATTAAGAAATGAAGCTG 840  
QY 841 gatgaagaacaaacaacacttgcgaacagagcagcagcagcagccttgcgcaagaagaag 900  
DB 841 GATGAAGAACAAACACACTTGGAAACAGAGACAGAGAGAGCAGCAGCTGGCCAGAGAGAG 900  
QY 901 tgcctgagacttaacagaactgctgcggcgcaatctcgaagcacaactgtcacctcacagaatc 960  
DB 901 TGCTTGAGACTTAACAGAACTGCTGGGCAATCTGAGACCAACCTGCACCTCACAGATCT 960  
QY 961 gaataagctcaactcagtcgaagaagaagaagttatcatatataataattgggaagaattcag 1020  
DB 961 GAAATAGCTCAACTCACTCAAGAAAAAGGTATATCATATATTAATTGGAAAAATTACAG 1020  
QY 1021 agaagaatlgaaagaatttgaagacagtgctccagcatlggaagagttaca tgaagcagtg 1080  
DB 1021 AGAAGAAATGAAGAAATTTGAGAGAAACAGTGTGCAGCATGGAGAGTACATGAGACGATG 1080  
QY 1081 aagcaaaagcttaagcagcgtcgtgataagcacaagccagcagcagcagcgtggtgag 1140  
DB 1081 AAGCAAAAGCTTAAGCGACGCTGATTAAGCACAGCAGCCACAGCCAGCGCTGTGTGCAG 1140  
QY 1141 ctctcagaagcagaagacagcgtctccctcgtgagagagcagagcgtgtcggagaaggtgagac 1200  
DB 1141 CTCTCTCAGCAAGCAGAGACAGCTTCTCTCTGAGAGGACAGAGCCTGTGTGGAAGAGTGGAC 1200  
QY 1201 cggctcggagcccaagttacccaagcatgccaacatctgattgctgcagcttga tgaagacaga 1260  
DB 1201 CGGCTGGGAGCCAGTGTACCCAGCATGSCCAATTCGATTCGTGACCTGTGATGGAACAGA 1260  
QY 1261 gtgaataaataatgaattacaagaagataattacatlcacatcgtgttgaacttaataatgcca 1320  
DB 1261 GTGAATAAATTAATGAATTAACAAGAAGATATTATTCATTCATCTGTTTGAAGCTTAATATGCCA 1320  
QY 1321 caacgacacagacgtctccagaggtgaacagcgtcagcgtcagcgtgaagtgagggtctgtccta 1380  
DB 1321 CAACGACACAGACGTCTCCAGGGTGACACCGCTCAGCTCAGTGAGGGGCTGTCTCTCA 1380



1381 tcaacgcgagcgtctgccccgacgacgtcggcgtgagctgagctctactctagctga 1440  
|||||  
1381 tcaacgcgagcgtctgccccgacgacgtcggcgtgagctgagctctactctagctga 1440  
|||||  
1441 gcaagctcctggtgtatgttttcaagaatgagcttgaagtatglttcaaatctgctcat 1500  
|||||  
1441 gcaagctcctggtgtatgttttcaagaatgagcttgaagtatglttcaaatctgctcat 1500  
|||||  
1501 tccgtatgtagtatacatatgattttcaataatgtaactttttaagaaga 1552  
|||||  
1501 tccgtatgtagtatacatatgattttcaataatgtaactttttaagaaga 1552  
|||||  
RESULT 2  
AF161348 1325 bp mRNA PRI 01-FEB-2000  
LOCUS AF161348 Homo sapiens HSPC085 mRNA, partial cds.  
DEFINITION AF161348  
ACCESSION AF161348  
VERSION AF161348.1 GI:6841109  
KEYWORDS FLI CDNA.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1325)  
Zhang, Q.H., Ye, M., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,  
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.  
Human partial CDS cloned from cd34+ stem cells  
Unpublished  
2 (bases 1 to 1325)  
Zhang, Q.H., Ye, M., Zhou, J., Shen, Y., Wu, X.Y., Guan, Z.Q., Wang, L.,  
Fan, H.Y., Mao, Y.F., Dai, M., Huang, Q.H., Chen, S.J. and Chen, Z.  
Direct Submission  
Submitted (14-MAY-1999) Shanghai Institute of Hematology, Shanghai  
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II,  
Shanghai 200025, P. R. China  
FEATURES  
SOURCE  
1. 1325  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CBDAND10"  
/cell\_type="cd34+ stem cell"  
/tissue\_type="cord blood"  
147..>1325  
/cdoon\_start=1  
/product="HSPC085"  
/protein\_id="AAP28908.1"  
/db\_xref="GI:6841110"  
/translation="MAKSPENSTLEELIGYORSIRREHASRSIHOTLCALKEGNYTIG  
EDAPNISFSTSVGNEADARTAPPELOQSHAVNQLDLROQDKSEVSPSRKMSPL  
RSLEHEFTNMTMDLVHTINDOSQYIHLEAEVKFCERLSGKMKIQVAVLENEGL  
QOQLKSORQETLEQFLDLSGNMHNSTITGDSGEGTSKRPFSHDNDFFKAAS  
AGEQLELEKLTLYEKECEIYESQLKFLRNDLARYORICEDLKEQLKEKLELLAANTC  
NRVGLCLCKCAHEAVLSQHTNMVQCTIEFLVKEPRDLMALVSVRSLSLDTQORRA  
SAYQVQVOVLQISEANFEKTKALIQCDQLKEERQAEERLEKELELAQOEKRALEKDM  
M"  
BASE COUNT 430 a 269 c 338 g 288 t  
ORIGIN  
Query Match 32.5%; Score 504; DB 51; Length 1325;  
Best Local Similarity 100.0%; Pred. NO. 5e-106;  
Matches 504; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

942 CATAAAGATTCTTCTGGCTGCTATACTGTGAACCGTGTTGGTCTTTGGAAA 1001  
|||||  
196 tgtgctcaagcagcagcgtgtcttcccaaacccactactaagttcaatgacacatc 255  
|||||  
1002 TGTGCTCAGACATGAAGCTGTCTTCCCAAAACCACTTAATGATATGACAGACATC 1061  
|||||  
256 gaaagactggttaagaagaagatgactgtgtctgcactgtttccgtaaggagcagc 315  
|||||  
1062 GAAAGACTGTTAAAGAAAGATGATGATGATGCTGCACTAGTTCCGTAAGAGACAC 1121  
|||||  
316 ttggcagatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 375  
|||||  
1122 TTGGCAGATACCCAGCAAGACAGCAAGCTGTATGACAGCTGAACAAATTTTGCA 1181  
|||||  
376 atatctgaggagcgaatttgaataaaccaagcgttcaatccagctgtgaccagttgag 435  
|||||  
1182 ATATCTGAGAGAGCCAAATTTGAAAAACCAAGGCTTTATTCAGTGTGACAGTTGAG 1241  
|||||  
436 aaggagctggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 495  
|||||  
1242 AAGGAGCTGGAGAGCGAGCGGACCTTGAAAAAGAACTTGACATTCAGCAGAGANA 1301  
|||||  
496 agggccattgagaagacatgatg 519  
|||||  
1302 AGGCCATTGAGAAAGACATGATG 1325  
|||||  
RESULT 3  
HSU79271/c 1198 bp mRNA PRI 25-MAR-1997  
LOCUS HSU79271/c Human clones 23920 and 23921 mRNA sequence.  
DEFINITION HSU79271  
ACCESSION HSU79271  
VERSION U79271.1 GI:1710237  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1198)  
Ricafrete, J.Y., Wentland, M.A., Morley, K.C., Muzny, D.M., Ding, Y., Liu, W.,  
Andersson, B., and Gibbs, R.A.  
A 'double adaptor' method for improved shotgun library construction  
Anal. Biochem. 236 (1), 107-113 (1996)  
JOURNAL 96207227  
MEDLINE  
REFERENCE 2 (bases 1 to 1198)  
Yu, W., Andersson, B., Morley, K.C., Muzny, D.M., Ding, Y., Liu, W.,  
Ricafrete, J.Y., Wentland, M.A., Lennon, G., and Gibbs, R.A.  
Large Scale Concatenation cDNA Sequencing  
Unpublished  
3 (bases 1 to 1198)  
Yu, W., and Gibbs, R.A.  
Direct Submission  
Submitted (22-NOV-1996) Molecular and Human Genetics, Baylor  
College of Medicine, One Baylor Plaza S930, Houston, TX 77030, USA  
FEATURES  
SOURCE  
1. 1198  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="23920, 23921"  
/sex="female"  
/tissue\_type="brain"  
/dev\_stage="infant"  
/clone\_lib="Soares library INIB from IMAGE consortium"  
BASE COUNT 380 a 216 c 250 g 352 t  
ORIGIN  
Query Match 21.0%; Score 326.4; DB 11; Length 1198;  
Best Local Similarity 99.4%; Pred. NO. 4.4e-65;  
Matches 338; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

|    |      |   |      |
|----|------|---|------|
| Db | 1152 | CCAGTTACCAGCATGGCCACAACTGATTTGGTGTGACCTGGATGGAAACGAGTGAATTAAT     | 109  |
| Qy | 1272 | gaattacaagaagatatttlaattcatctgtgtttaagacttaataatgcacaacgcaacc     | 133  |
| Db | 1092 | GATTTCACAAAGATATTTCATATTCCTGGTTTAAAGCTTAATATGGCAACAGCACAC         | 103  |
| Qy | 1332 | gacctcccaaggagacacccgcctcagccttgcagctggagctgtgtcctcatcaacgcgggc   | 139  |
| Db | 1092 | GACCTTCACAGGAGGAGACACGGCCTCACGCTGACATGGGGCTGGTCCCTCATCAACGGGGC    | 973  |
| Qy | 1392 | gctgtcccccagacgcagtcgggcttgcagcttgcagcttgcacttgccttgcagcaga-ctcct | 145  |
| Db | 972  | GCTGTCCCCGACACGACGCGGGCTGGAGCTGGAGCTTCGACTCTAAGCGACAAAGCTCCT      | 913  |
| Qy | 1451 | ggctatgcttttcagaaatgctgcgaagtatgctgtttaaactcgtcctatctgtatgcta     | 151  |
| Db | 912  | GGTGATGTTTTTCAGAAATGGCTTGAAGTTATGTTTAAATCGTCGATCTGTATGCTTA        | 853  |
| Qy | 1511 | ggtatatacatatgatttccaataatgaacttcttaaga                           | 1550 |
| Db | 852  | GGTTATACATATGATTTTCAATAATCAACTTTTAAAGA                            | 813  |

| RESULT     | 4   |
|------------|---|
| G27428     | 576 bp                                      |
| LOCUS      | DNA   |
| DEFINITION | human STS SHGC-31921, sequence tagged site. |
|            | 28-JUN-1996                                 |

|          |          |  |
|----------|----------|--|
| VERSION  | G27428.1 | GI:1396147                             |
| KEYWORDS | STS; STS | sequence; primer; sequence tagged site |
| SOURCE   | human.   |  |

REFERENCE  
1 (bases 1 to 576)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

JOURNAL  
COMMENT  
unpublished (1996)

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 415/7259687  
Fax: 415/7259689  
Email: myersrshgc.stanford.edu

Primer A: TATTTTTCAGTTACCCAGCA  
Primer B: TCGGTGTGGCATATTAGTC  
STS size: 127  
PCR Profile:

|                     |                             |
|---------------------|-----------------------------|
| Initial incubation: | 94 degrees C for 90 seconds |
| Denaturation:       | 94 degrees C for 15 seconds |
| Annealing:          | 62 degrees C for 23 seconds |
| Polymerization:     | 72 degrees C for 30 seconds |
| PCR Cycles:         | 30                          |
| Thermal Cycler:     | Perkin Elmer 9600           |
| Protocol:           |                             |

|                 |                     |
|-----------------|---------------------|
| Template:       | 25 ng               |
| Primer:         | each 1 $\mu$ M      |
| dNTPs:          | each 200 $\mu$ M    |
| Tag Polymerase: | 0.05 units/ $\mu$ l |
| Total Vol:      | 10 $\mu$ l          |

|                     |     |    |
|---------------------|-----|----|
| MgCl <sub>2</sub> : | 2.5 | mm |
| KCl:                | 50  | mm |
| Tris-HCl:           | 20  | mm |
| pH:                 | 8.3 |    |

-- Prepared with primer pairs provided by Sandoz, derived from R540922  
-- Washington University/Merck EST sequence.

## FEATURES

```

source      1. .576      /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="1"
STS         24. .150
primer_bind 24. .45
primer_bind complementary(130. .150)
BASE COUNT 144 a      123 c      111 g      190 t
ORIGIN

```

|                       |                 |                    |           |             |
|-----------------------|-----------------|--------------------|-----------|-------------|
| Query Match           | 16.6%;          | Score 257.2;       | DB 13;    | Length 576; |
| Best Local Similarity | 95.1%;          | Pred. No. 3.8e-49; |           |             |
| Matches 330;          | Conservative 0; | Mismatches 9;      | Indels 8; | Gaps 6;     |

Qy 1212 ccagttaccacgcat-gccacaatctgtatttgcacct-  
ggatgygaacgagtgaataaa 12  
Db 32 CCAGTTACCACGACATNGCACAACTGTATTGCTGACTGGGATGGGAACAAGAGTCAATAA 91

Qy 1270 atgaattacaaagagatatattacatcactc -gtttaagacttaataatgcccacaagcac 132

Db 92 ATGATTTACAAAGACATATTTCATTCATCTCGGGTTCAGACTTAATAATGCCACAACGCAC 151

QY 1329 cagcagcttccca-gggtgacacccgcctcagctcagtggtgtgtccataaacgc 1330  
|||||  
152 CACGACCTTCCAGGGGTGACACAGGCTCAGCCTGCAGTGGGGCTGNNCTCATCAACGC 211

```
Oy      1388 ggggagctgtcccgacgcagcagtctgsg--ctygagcttgaatctgacttagtcyagcacaa   1445
```

(The following sequence continues from the previous block)

```
Db       212 GGGGCGTGTCCCGACGCAGTCGGGGCTGGGAGTGAGACTTTAAGCTGAGCAGA    271
```

```

QY 1446 --ctccctggtgtaagtttttcagaataagcgctggaagtatgctgtttaatcctgcattcg 150
      + |||||
Db 272 GCTCCTGGGTATGTTTTCAGAANAAGCGTGAAGTATGTGTTAAATCTGCTCATTCG 331

```

```
OY      1504 catgctaggttatacatatgatlttcataataatgaactttttaaga 1550
          |||||
Db       332 TATGCTAGGTATACATATGATTTCAATAAATGAACTTTAAAGA 378
```

|            |                                     |
|------------|-------------------------------------|
| RESULT     | 5                                   |
| LOCUS      | 166494/c                            |
| DEFINITION | Sequence 14 from patent US 5670367. |
| DNA        | 7218 bp                             |
| PAT        | 28-DEC-1997                         |

|          |          |            |
|----------|----------|------------|
| VERSION  | I66494.1 | GI:2724471 |
| KEYWORDS | .        |            |
| SOURCE   | Unknown. |            |

Unclassified.  
1 (bases 1 to 7218)  
Dorner, F., Scheifflinger, F. and Falkner, F. Gunter

|          |                                     |
|----------|-------------------------------------|
| JOURNAL  | Patent: US 5670367-A 14 23-SEP-1997 |
| FEATURES | Location/Qualifiers                 |
| source   | 1..7218                             |

|            | /organism="unknown" |        |        |        |            |
|------------|---------------------|--------|--------|--------|------------|
| BASE COUNT | 1944 a              | 1491 c | 1486 g | 1929 t | 368 others |
| ORIGIN     |                     |        |        |        |            |

|                       |                   |                    |           |              |
|-----------------------|-------------------|--------------------|-----------|--------------|
| Query Match           | 5.8%;             | Score 90.6;        | DB 5;     | Length 7218; |
| Best Local Similarity | 6.9%;             | Pred. No. 1.1e-10; |           |              |
| Matches 30;           | Conservative 252; | Mismatches 151;    | Indels 0; | Gaps 0;      |

Qy 421 tctgaccagtttgaggagagcgtgagagcgagcgacttgaanaagaattgca 48  
||| || | ||| | ||| | | ||| |||  
Db 1496 TGTAGGCATCATCTGTAATTACCACTACTATGCCAGTAGTAAAGAAGATTCAGGAATTTGGT 14

Qy 481 ttcagcaagaaaggccattgaaacatgtgaaaaagaataaacgaagaa 540  
| :: :::::::::::::: : ::::: : : :::::::::::::: : :::::  
Db 1436 acrrr 1377

[illegible]

|            |   |
|------------|---|
| FT         | /db_xref="taxon:32644"  |
| FT         | /organism="unidentified"  |
| xx         |   |
| SQ         | Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;  |
|            |   |
|            | Query Match      4.3%; Score 67.2; DB 24; Length 1686;<br>Best Local Similarity    45.1%; Pred. No. 2,5e-05;<br>Matches 249; Conservative    0; Mismatches 303; Indels    0; Gaps    0;   |
| OY         | 328 cagcaagaagaaagcgaagtgcctatgtaacaagtgtaacaaagtttgcataatatctgaagaa 387<br>Db    71 CAATTAATAAATAATTAAAAATGTAGAAAAAGAAGAAAGAAAATGAAAGAAATATAAAAAAGAG 130   |
| OY         | 388 gccaatcttgaaaaaaccaaggcctttaaccagtgtagccagttlagaagaagagctggag 447<br>Db    131 AACCAAGAAAGACAGAACACAGACAAGAACAGAACAGAACAGAAATCGTAGAAGAACAG 190  |
| OY         | 448 aggcagcgcygagcgacttgaaaagaacctgtcatctcagcaagaagaaagggccactgg 507<br>Db    191 AACCAAGATGAAGAAAGAACAGAACAGAGGAAAGATGAGAAAGAAAGAAAGAAAGAAAGAG 250   |
| OY         | 508 aaagacatgatgtaaaagaaataaaccgaagaagaagggagtcattggaatcaaagatgtg 567<br>Db    251 AAGAAGAAAGAACAGAACAGAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCAGATGTAG 310   |
| OY         | 568 atcttgtctccaagaattctgcccaacttgtagagcccagcggtggaaaaggttaacaagaagaa 627<br>Db    311 AAGAAAGAAAGAACAGAACAGAACAGATGTAGAAAGAAAGAAAGAAAGAAAGATGAGAGAG 370  |
| OY         | 628 attcaagcttattaatcaacttgagaggaataatcaaaagccagctgtgctctcggaaatggat 687<br>Db    371 AAGAAGAAAGAAAGAAACAAATGTAGAAAGAAACAAATGTAGAGATGAAACAAATGAGATGAG 430   |
| OY         | 688 gtcacaagaagtgtgtgaggaatcgcgtcatacagctgaataaaccacatgagagaagat 747<br>Db    431 AAAATGAAGAGAACAGAACAGAACAGAAAGAAAGAAAGAAACAAAGCATGAGTGAAGAGAAC 490  |
| OY         | 748 gaggaagaagaagagcacagagagfttcgagcgcaaaactcaagatcagaggtcttggaaatraa 807<br>Db    491 AAGATGAAGAGAACAGAACAGTGAAGAAAGAAAGAAAGAAAGAAACAGAACAGTGTAGAG 550   |
| OY         | 808 gatcaggaataatagaaaattgagaatagaaactgtagtgaagaagcaacaacacttgyaagac 867<br>Db    551 AACCAAGAAAGACTATATCTGTAAAAAGAAATGTAGAGATGAGAGAAAAAAGAAAAAGAG 610  |
| OY         | 868 gaggcagcagaag 879<br>Db    611 AAGAACAAGAGAG 622  |
| RESULT     | 7   |
| LOCUS      | AC007461/c DNA PRI 04-JUN-1999  |
| DEFINITION | Homo sapiens chromosome 17, clone 84_E_24, complete sequence.   |
| ACCESSION  | AC007461  |
| VERSION    | AC007461.8 GI:4996920   |
| KEYWORDS   | HTG.  |
| SOURCE     | human.  |
| ORGANISM   | Homo sapiens  |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;<br>Ethelera; Primates; Catarrhini; Homnidae; Homo.<br>1 (bases 1 to 180385)<br>Britten,B., Linton,L., Nusbaum,C. and Lander,E.<br>Unpublished<br>2 (bases 1 to 180385)<br>Baker,J., Baldwin,T., Barna,N., Beckery,R., Benn,J., Brown,A.,<br>Castle,A., Cerny,J., Colangelo,M., Collins,S., Collimore,A.,<br>Cook,P., DeRellano,K., Depayre,E., Devon,K., Dewar,K.,<br>Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,<br>Funke,R., Gaee,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G., |
| AUTHORS    |   |

Hagos, B., Hearford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., McGowan, N., McGowan, P., McGurk, A., McKernan, K., McLaughlin, D., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

Direct Submission  
Submitted (04-MAY-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 180385)

REFERENCE  
Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, D., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearollano, K., Depayre, E., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Hearford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McGowan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE  
JOURNAL  
Submitted (04-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 4, 1999 this sequence version replaced gi:4966400.  
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT

Only 180385 base pairs from the middle of this clone are being submitted. The remainder overlaps either accession number AC005152 (WICGR project L335) or accession number AC005243 (WICGR project L343).

FEATURES  
source  
location/Qualifiers  
1.180385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="84.E.24"  
/clone\_1lb="Alan Buckler -- per comm"  
/map="17"  
/chromosome="17"  
446..753  
/rpt\_family="AlusC"  
2375..2419  
/rpt\_family="(TTTGG)n"  
complement(2985..3103)  
/rpt\_family="L2"  
complement(3146..3352)  
/rpt\_family="L2"  
complement(3531..3639)  
/rpt\_family="L2"  
complement(3798..4394)  
/rpt\_family="L2"  
5531..5631  
/rpt\_family="MER81"  
5648..5674  
/rpt\_family="(CAAAA)n"  
5812..6020  
/rpt\_family="L1MC4"  
6180..6224  
/rpt\_family="AT-rich"  
6303..6611  
/rpt\_family="AlusP"  
7378..7618  
/rpt\_family="MIR"  
9593..9623

/rpt\_family="(GGCA)n"  
complement(10666..10923)  
/rpt\_family="AlusG"  
10958..11088  
/rpt\_family="MIR"  
11752..11914  
/rpt\_family="L2"  
complement(11915..12219)  
/rpt\_family="AlusX"  
12220..12347  
/rpt\_family="L2"  
13110..13182  
/rpt\_family="GA-rich"  
13526..13564  
/rpt\_family="(TCC)n"  
13575..13660  
/rpt\_family="(TA)n"  
complement(14380..14658)  
/rpt\_family="AlusO"  
14742..14830  
/rpt\_family="(TA)n"  
14860..14967  
/rpt\_family="(TA)n"  
complement(14977..15122)  
/rpt\_family="L2"  
complement(15292..15385)  
/rpt\_family="L2"  
16672..16781  
/rpt\_family="L2"  
16802..16846  
/rpt\_family="(TGA)n"  
17225..17245  
/rpt\_family="(A)n"  
18411..18720  
/rpt\_family="AluB"  
complement(20715..20788)  
/rpt\_family="MLT1J2"  
complement(20822..20916)  
/rpt\_family="MLT1J2"  
20956..21137  
/rpt\_family="MIR"  
21830..21872  
/rpt\_family="AT-rich"  
21957..22257  
/rpt\_family="AlusX"  
complement(22438..22538)  
/rpt\_family="MLT1J"  
complement(23375..23667)  
/rpt\_family="AluY"  
24553..24858  
/rpt\_family="AluY"  
complement(25021..25234)  
/rpt\_family="MIR"  
26957..27261  
/rpt\_family="AluY"  
complement(27524..27669)  
/rpt\_family="MLT1B"  
complement(27670..27964)  
/rpt\_family="AlusP"  
complement(27965..28210)  
/rpt\_family="MLT1B"  
complement(28228..28522)  
/rpt\_family="AlusG"  
complement(28592..28700)  
/rpt\_family="MIR"  
complement(29353..29537)  
/rpt\_family="MIR"  
29802..30095  
/rpt\_family="AlusG"  
complement(30099..30219)  
/rpt\_family="MIR"  
30227..30267  
/rpt\_family="(TG)n"

```

Repeat_region      30326. .30321
/rpt_family="CA)n"
complement(30322. .30611)
Repeat_region      /rpt_family="AluBc"
complement(30956. .31151)
Repeat_region      /rpt_family="MER3"
complement(31533. .31633)
/rpt_family="L2"
Repeat_region      31810. .31858
/rpt_family="L2"
Repeat_region      32028. .32107
/rpt_family="L1PA4"
complement(32919. .33182)
Repeat_region      /rpt_family="MIR"
34169. .34314
Repeat_region      /rpt_family="CT-rich"
34774. .35007
Repeat_region      /rpt_family="MIR"
35589. .35621
Repeat_region      /rpt_family="AT-rich"
35763. .36142
Repeat_region      /rpt_family="MER41A"
36143. .36452
Repeat_region      /rpt_family="AluSg"
36453. .36636
Repeat_region      /rpt_family="MER41A"
38011. .38071
Repeat_region      /rpt_family="CT-rich"
38083. .38362
Repeat_region      /rpt_family="(TTC)n"
38235. .38319
Repeat_region      /rpt_family="(TTC)n"
38320. .38499
Repeat_region      /rpt_family="(TTC)n"

```

|                           |        |                    |           |                |
|---------------------------|--------|--------------------|-----------|----------------|
| Query Match               | 4.28;  | Score 65.8;        | DB 40;    | Length 180385; |
| Best Local Similarity     | 45.28; | Pred. No. 6.2e-05; |           |                |
| Matches 241; Conservative | 0;     | Mismatches 292;    | Indels 0; | Gaps 0;        |

|    |       |   |       |
|----|-------|---|-------|
| OY | 347   | ctltgaaacaggctgaataaagaatttgcataatctcgggaagccaatcttgaataaaca    | 406   |
|    |       |   |       |
| Db | 38635 | CATATTCAAACTGTGAGTGTGTGCAGAAATGCAAAATGAAAGAAAGAAAGAA            | 38576 |
| OY | 407   | agccttaatccagctgtgtaccagtttgagaagaagctggaagcgagcgccagcttg       | 466   |
|    |       |   |       |
| Db | 38575 | AGGAAAGAGAGAGGAGGAGGAGGAAGAAAGAAAGAAAGAGAGGAGGAAGAGGAGC         | 38516 |
| OY | 467   | aaaagaaccttgcatctcagcaagaagaagggccattggaagaagacatgatgaaaaag     | 526   |
|    |       |   |       |
| Db | 38515 | AGGAAGAGAGAGAGGAGGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGG         | 38456 |
| OY | 527   | aataacgaagaagaaggaagatcactggatcctaagaatgttgatcttgcctcgaatatgt   | 586   |
|    |       |   |       |
| Db | 38455 | AGAAAGCAAGAGAGAGAGGAAGGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA    | 38396 |
| OY | 587   | cccaactggagcccaagctgggaagaaggttacaagaagaagaattcactaatactaac     | 646   |
|    |       |   |       |
| Db | 38395 | AGGAAGAGAGAGAGGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA         | 38336 |
| OY | 647   | tggaggaattccaagccagcttgctctcgggaattgatgtcacaaggtgtgtgag         | 706   |
|    |       |   |       |
| Db | 38335 | AGGAAGAAAGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG       | 38276 |
| OY | 707   | aatgtgcgtatccagcttgatataaaccacaatcgggaagaagatggagcagaagaagcgaca | 766   |
|    |       |   |       |
| Db | 38275 | AGGAGAGAGAGAGGAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGA      | 38216 |
| OY | 767   | gagagttcagagcaaaaactacaagagatcttgaattcaagaatcaggaataatagaagaat  | 826   |
|    |       |   |       |
| Db | 38215 | AGGAGAGAGAGGAGGAGGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAAGG     | 38156 |
| OY | 827   | tgagaatagaacttgatgaaagcaacaacacttggaaacagagcagcagaag            | 879   |
|    |       |   |       |

| RESULT     | 8  |               |     |     |             |
|------------|--|---------------|-----|-----|-------------|
| CEL33368   |  |               |     |     |             |
| LOCUS      | CEL33368   | 39103 bp      | DNA | INV | 08-JUL-1998 |
| DEFINITION | Caenorhabditis elegans   | cosmid C3368. |     |     |             |
| ACCESSION  | U5154  |               |     |     |             |
| VERSION    | U5154.1  | GI:1255414    |     |     |             |
| KEYWORDS   |  |               |     |     |             |
| SOURCE     | Caenorhabditis elegans.  |               |     |     |             |
| ORGANISM   | Caenorhabditis elegans   |               |     |     |             |
| REFERENCE  | Eukaryotic Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  |               |     |     |             |
| AUTHORS    | 1 (bases 1 to 39103)<br>Wilson, R., Alnsough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Showkeen, R., Smaldon, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vautin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wohlman, P. |               |     |     |             |
| TITLE      | 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans   |               |     |     |             |
| JOURNAL    | Nature 368 (6466), 32-38 (1994)  |               |     |     |             |
| MEDLINE    | 94150718   |               |     |     |             |
| REFERENCE  | 2 (bases 1 to 39103)<br>Gatlung, S. and Wu, X.   |               |     |     |             |
| AUTHORS    | The sequence of C. elegans cosmid C3368  |               |     |     |             |
| JOURNAL    | unpublished (1998)   |               |     |     |             |
| REFERENCE  | 3 (bases 1 to 39103)<br>Waterston, R.  |               |     |     |             |
| AUTHORS    | Direct Submission  |               |     |     |             |
| TITLE      | Submitted (29-MAR-1996)  |               |     |     |             |
| JOURNAL    | 4 (bases 1 to 39103)   |               |     |     |             |
| REFERENCE  | Robert Waterston   |               |     |     |             |
| AUTHORS    | Waterston, R.  |               |     |     |             |
| TITLE      | Direct Submission  |               |     |     |             |
| JOURNAL    | Submitted (08-JUL-1998)  |               |     |     |             |
| REFERENCE  | Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA   |               |     |     |             |
| AUTHORS    | Submitted by:  |               |     |     |             |
| COMMENT    |  |               |     |     |             |

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is T07H8, 200 bp overlap; 3' cosmid is F20A1, 2000 bp overlap. Actual start of this cosmid is at base position 1 of CELC3368 ; actual end is at 6614 of CELF20A1

NOTES:



|    |       |   |       |
|----|-------|---|-------|
| Db | 31546 | ATGATGACGAGCAAAAAGCAAAAGTAAAGAAAAGACAGACAGAAAAGGACGATGATG           | 31605 |
| QY | 509   | aagcattatgtaaaaaaggaataacagaaagaaggagctacatggatccaagaatglttga       | 568   |
| Db | 31606 | ACGATGAAGATGAAAAATGATTAAGAAAAAGATTAACAAAAAAAAGCAAGATGCAAGATG        | 31665 |
| QY | 569   | tctgtctcagaatactcccaactcggagggcccgatgtaaaagtctcaaaagaaaaga          | 628   |
| Db | 31666 | ATGATGACAAATGAGGATGACAAAAAAGAGGACAAAAAATAAAGG--AAGGATGTTAAGG        | 31722 |
| QY | 629   | ttcagcttatcaactcggagaagaattcaagccagctgctcttcggaaatgagt              | 688   |
| Db | 31723 | AAAAAGATGATGATGATGATGAGGAGGAGCAAGCATTAACAAAAGAAAGAACACAAAAAGAAAGATG | 31782 |
| QY | 689   | tcacaagaagtgtgtgagaaatgcgctctcagctgaataaaacacactgagaagaagt          | 748   |
| Db | 31783 | ACGACGACGAGGAAGATTAATAAAGAAAGACAAAGAAAAAGAAAAAGCAGCATGATGACGATG     | 31842 |
| QY | 749   | aggcagaaaaaggacacagagagttcagagcaaaaaactaacacaggaactcttgaattlaag     | 808   |
| Db | 31843 | AGGATGAGCATTAATTAAGAAAAAAGGATTAAGAAAAAGACAGAGGATGATTAAGGATGATGAAG   | 31902 |
| QY | 809   | atcaggaatacagagaataattgagaatagaacttggatgtaagaacaaacaacacttggaaacsg  | 868   |
| Db | 31903 | ATGAGCAAAAAGCAGACAGACAAAGAAAGAAAGGATTAACAAAAAAGATGATGACGATGACG      | 31962 |
| QY | 869   | agcagcgcaagaag  | 880   |
| Db | 31963 | ATGATGAAAAAAG   | 31974 |

|            |  |
|------------|--|
| RESULT     | 9  |
| AC006884/c |  |
| LOCUS      | AC006884 193188 bp DNA HTG 26-FEB-1999   |
| DEFINITION | Ceenorhabditis elegans clone Y57E12, *** SEQUENCING IN PROGRESS  |
| ACCESSION  | ***, 4 unordered pieces.   |
| VERSION    | AC006884.2 GI:4309911  |
| KEYWORDS   | HTG; HTGS_PHASE1.  |
| SOURCE     | Ceenorhabditis elegans.  |
| ORGANISM   | Eukaryotes; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditidae;<br>Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Ceenorhabditis<br>1 (bases 1 to 193188) |
| REFERENCE  | Waterston,R.H.   |
| AUTHORS    | The sequence of Ceenorhabditis elegans clone   |
| TITLE      | Unpublished  |
| JOURNAL    | 2 (bases 1 to 193188) -  |
| REFERENCE  | Waterston,R.H.   |
| AUTHORS    | Direct Submission  |
| TITLE      | Submitted (24-FEB-1999) Genome Sequencing Center, Washington   |
| JOURNAL    |  |

```

COMMENT
* Mar 1, 1999 this sequence version replaced gi:1263464.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 2784: contig of 2784 bp in length
*
* 2785 2793: gap of unknown length
*
* 2800 37804: contig of 35005 bp in length
*
* 37805 37819: gap of unknown length
*
* 37820 74165: contig of 36346 bp in length
*
* 74166 74180: gap of unknown length
*
* 74181 193188: contig of 119008 bp in length.
*
* Location/Qualifiers
*
* 1..193188
* /organism="Caenorhabditis elegans"

```

|            |         |                       |  |
|------------|---------|-----------------------|--|
| BASE COUNT | 63610 a | /db_xref="taxon:6239" |  |
| ORIGIN     | 33574 c | /clone="Y5E12"        |  |
|            | 33884 g |                       |  |
|            | 62075 t |                       |  |
|            | 45      | others                |  |

|                       |              |                    |                 |                |
|-----------------------|--------------|--------------------|-----------------|----------------|
| Query Match           | 4.0%;        | Score 62.4;        | DB 41;          | Length 193188; |
| Best Local Similarity | 45.8%;       | Pred. No. 0.00038; |                 |                |
| Matches 253;          | Conservative | 0;                 | Mismatches 296; | Indels 3;      |
|                       |              |                    |                 | Gaps 1;        |

[illegible]

|              |   |
|--------------|---|
| RESULT       | 10  |
| MUSMHC4H2S/C |   |
| LOCUS        | MUSMHC4H2S 17522 bp DNA ROD 02-AUG-1999   |
| DEFINITION   | Mus musculus complement component C4A (C4A) gene, complete cds  |
| ACCESSION    | M17440 J050905  |
| VERSION      | M17440.1 GI:199279  |
| KEYWORDS     | B1 repeat; complement component C4.   |
| SOURCE       | house mouse.  |
| ORGANISM     | Mus musculus  |
| REFERENCE    | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;<br>Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.                                 |
| AUTHORS      | 1 (bases 1 to 1502)   |
| TITLE        | Rosa, P.A., Sepich,D.S., Robins,D.M. and Ogata,R.T.<br>Constitutive expression of SLP genes in mouse strain B10.WR<br>directed by C4 regulatory sequences |
| JOURNAL      | J. Immunol. 139, 1568-1577 (1987)   |
| MEDLINE      | 87309760  |
| REFERENCE    | 2 (bases 1 to 17522)  |
| AUTHORS      | Ogata,R.T., Rosa,P.A. and Zepf,N.E.   |
| TITLE        | Sequence of the gene for murine complement component C4   |
| JOURNAL      | J. Biol. Chem. 264 (28), 16565-16572 (1989)   |
| MEDLINE      | 89380278  |

COMMENT Draft entry and computer-readable sequence for [1] kindly submitted  
by R.T. Ogata, 12-SEP-1989.  
FEATURES Location/Qualifiers  
SOURCE 1. 17522  
/organism="Mus musculus"  
/strain="B10.NR"  
/db\_xref="taxon:10090"  
/chromosome="17"  
/tissue\_type="liver"  
mRNA join(1443..1564,1665..1857,2314..2515,2724..2794,  
2874..2962,3256..3338,3439..3535,4051..4156,4232..4364,  
4688..4803,4889..5065,5211..5396,5492..5677,6331..6489,  
7579..7705,8375..8449,8703..8897,8990..9104,9209..9279,  
9469..9608,9737..9940,10416..10467,10554..10643,  
10769..10978,11137..11212,11316..11472,11568..11684,  
11773..11944,12139..12371,12453..12620,13115..13174,  
14175..14268,14354..14534,14629..14719,14849..14923,  
15437..15539,15675..15764,15842..15940,16075..16158,  
16443..16575,16713..16953)  
/gene="C4a"  
/product="complement component C4a"  
1443..16953  
gene /gene="C4a"  
1443..1564  
exon /gene="C4a"  
1500..1556  
sig\_peptide /number=1  
1500..1556  
CDS /gene="C4a"  
join(1500..1564,1665..1857,2314..2515,2724..2794,  
2874..2962,3256..3338,3439..3535,4051..4156,4232..4364,  
4688..4803,4889..5065,5211..5396,5492..5677,6331..6489,  
7579..7705,8375..8449,8703..8897,8990..9104,9209..9279,  
9469..9608,9737..9940,10416..10467,10554..10643,  
10769..10978,11137..11212,11316..11472,11568..11684,  
11773..11944,12139..12371,12453..12620,13115..13174,  
14175..14268,14354..14534,14629..14719,14849..14923,  
15437..15539,15675..15764,15842..15940,16075..16158,  
16443..16575,16713..16854)  
/gene="C4a"  
/codon\_start=1  
/product="complement component C4a"  
/protein\_id="AA39561.1"  
/db\_xref="GI:387440"  
/translation="MRLMGLAWVFSFCASSLQKPRLLFSPVNLGTPLSVQVL  
DAPGGEVKGVSFLRNPKGSCSPKDKDFVLLSEVLEEDVRSQGLDRLR  
APHTQVAOSPLMNTAFKATNGGVNLLFSSRGHLEVOTDPIVPCRGVRRRA  
LDQKMRSTDFLTITVENSHQRLVKKKEIFSTISIFODAFITDIPISERGKTSARS  
DGLESNSTHEVKKIVLPNFEVKITPMKPTILWVPSNSEIQIDQARYICRPVG  
VAYTRFALMDEQGRFVLRGLETAQKLVGEKTHISIKDQFQALDKINIGVRLDGL  
RLVAAVAVIESPGGEMEAELTSWRPVSASFSLDSTRKHLVGAHFLQALVQESL  
GSEASNPVKYSATLVGSDSOVLDIOOSTNGIGOVSLSPPIPTVELRLVAGSL  
YPATRLTVQAPPSRGVGLSLIEPLDRSPVGPFTLINOPVGIAPATSGHYVYMT  
SRQIYAMGRPRKTYTVSYLVNDQALPSPYPAITYHGHGHPYANSLNINOSRDE  
GKIDLVDAKEYRNADMKRLIQTDKALVALAVTALYAVGRSHKPLDKSKE  
VINSYVCGGGGDALQVFDAGLAFSDDBRLTQREDLSCKREKSRQRANFQ  
KAVSEKIGQYSSPPAKRCQDGMKLPKPKRCEQARAVPQACREPLSCCKPVE  
RRNQTROAHLLARNHMLDEEDLIDEDDILVFPSPENMLMREVPDSSLLVWLP  
DSMTTWEIHGVSLSKSGICVAKPTRVKPEFLHLRLPISIRFEQFLRPVLVY  
LNDVAVSYHYTPYEGICLAGGMAQOYVPASARPAFVSVPYPAANVPLKYAR  
GVPLDQAVSKTIQIEKGAITHRELVYNDPLNLDLRTIESSDPTNVDDSES  
LVRYTASEPLETSGSEGLSPGVASLLRLPQGCAGETMTYLAFTLASNLDTDEW  
SKLSPETKHAVIDLIQGYMRIQFRRNDGSGFAMLRHDSWTWLTAFVLILSLAQD  
VGNSEPKIETASWLIAQOLGDSFHPDPCPIYHBMOGGSLGDEVALAFVYIAH  
HGLDVPODDAKOLKRNVPASITKANSFLGKAGLIGAAHAAITVAALITMSAD  
LRNVAHNSLMAAETSEHLYWGLVLSQDKVYLRPTAPSPTEPVPQALMTEETTA  
YALHLHLRREGKMAADKASWLTGHQSFHAFSTQDVTYVTTDALSAVYASTHTE  
KALKVTLSSMGRNLKTHGLNLNNHGVKGLEELKFSLGSTISVYKVGNSKGTLLIKR  
TYNVLDKNTTCODLIQEVKTVGAVEYAMDANEDEYIDMAPDPSVLPQPTPIQ  
LEGRSRRRREARVAEODESRVOYTCVWRNKLGLSGAIDLITLLSGFHALRD  
LEKTLSSDRVSHFETDPHVILYFDVSPTPRCVGASQEVVGLVPPSSAVLD  
YSPDHKCSVPYAPTKSQLLATLCSGDVQCAKGRPKRLRSIRREVEDKGRMFE  
ACTYPRVEYFTVAVLNKEDGAAFLRESKITVQLHRRKDTMASIGQTRNPLASCR  
LRLPNKYEYLTMGDGETDNKGDPOYLIDSNWNTWIEBMPSEDMCKSTRHRAACPOLAD

mat\_peptide FLMEFSRGQV"  
join(1557..1564,1665..1857,2314..2515,2724..2794,  
2874..2962,3256..3338,3439..3535,4051..4156,4232..4364,  
4688..4803,4889..5065,5211..5396,5492..5677,6331..6489,  
7579..7705,8375..8403)  
/gene="C4a"  
/product="complement component C4 beta chain"  
1565..1664  
intron /number="C4a"  
/number=1  
1665..1857  
exon /gene="C4a"  
/number=2  
1858..2313  
intron /gene="C4a"  
/number=2  
2314..2515  
exon /gene="C4a"  
/number=3  
2516..2723  
intron /gene="C4a"  
/number=3  
2724..2794  
exon /gene="C4a"  
/number=4  
2795..2873  
intron /gene="C4a"  
/number=4  
2874..2962  
exon /gene="C4a"  
/number=5  
2963..3255  
intron /gene="C4a"  
/number=5  
3256..3338  
exon /gene="C4a"  
/number=6  
3339..3438  
intron /gene="C4a"  
/note="7"  
/number=6  
3439..3535  
exon /gene="C4a"  
/number=7  
3536..4050  
intron /gene="C4a"  
/note="comp1"  
/number=7  
3638..3760  
repeat\_region /rpt\_family="B1 repeat"  
4051..4156  
exon /gene="C4a"  
/number=8  
4157..4231  
intron /gene="C4a"  
/number=8  
4232..4364  
exon /gene="C4a"  
/number=9  
4365..4687  
intron /gene="C4a"  
/number=9  
4688..4803  
exon /gene="C4a"  
/number=10  
4804..4888  
intron /gene="C4a"  
/number=10  
4889..5065  
exon /gene="C4a"  
/number=11  
5066..5210  
intron /gene="C4a"



|               | Query Match   | Best Local Similarity | Matches | 3.8%;<br>45.9%;<br>238; Conservative | Score 59;<br>Pred. No. 0.0021;<br>0; Mismatches | DB 12;<br>Length 17522;<br>280; Indels | 1;<br>Gaps |
|---------------|---|-----------------------|---------|--------------------------------------|---|--|------------|
| exon          | /number=11<br>5211.. 5396<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| intron        | /number=12<br>5397.. 5491<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| exon          | /number=12<br>5492.. 5677<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| intron        | /number=13<br>5678.. 6330<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| exon          | /number=13<br>6331.. 6489<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| intron        | /number=14<br>6490.. 7578<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| exon          | /number=14<br>7579.. 7705<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| intron        | /number=15<br>7706.. 8374<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| repeat_region | /number=15<br>7869.. 8266<br>/rpl_family="MT repeat"  |                       |         |                                      |   |  |            |
| exon          | /number=16<br>8375.. 8449<br>/gene="C4A"  |                       |         |                                      |   |  |            |
| mat_peptide   | /number=16<br>9011(8416.. 8449, 8703.. 8897, 8990.. 9104, 9209.. 9279,<br>9465.. 9608, 9737.. 9940, 10416.. 10467, 10554.. 10643,<br>10769.. 10978, 11337.. 11212, 11316.. 11472, 11568.. 11684,<br>11773.. 11944, 12139.. 12371, 12453.. 12620, 13115.. 13174, |                       |         |                                      |   |  |            |
| QY            | 382 gaggaagcccaatttggaaaaaaccaagagcttcaatccagtgagccagttggaagagag 441  |                       |         |                                      |   |  |            |
| Db            | 7301 GAGGAGGAAGAAAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7242   |                       |         |                                      |   |  |            |
| QY            | 442 ctggaagagcagcgagagcgacttgaaaaagaacttgcattccagcaagagaaaaggcc 501   |                       |         |                                      |   |  |            |
| Db            | 7241 TAGTACGAAGAGAGAGAGAGAGAGAGATATGAAAGAAAGAAAGAAATATGAGAAAGAGAA 7182  |                       |         |                                      |   |  |            |
| QY            | 502 attgagaaagacatgataaagaagaataacgaaagaagaggtacatgagatacaag 561  |                       |         |                                      |   |  |            |
| Db            | 7181 GAAGAGGAGAGAGAGATGACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 7122  |                       |         |                                      |   |  |            |
| QY            | 562 atgtgtactctgtctcagaaatattgcccaactgagagcccaagtgtagaaaaggttaacaag 621   |                       |         |                                      |   |  |            |
| Db            | 7121 GAAGAGAGCAAAATTGAAAG 7062  |                       |         |                                      |   |  |            |
| QY            | 622 gaaaagatttcagctattcaatcaacttgaaggaattcaaacagccttggtcttcgagaa 681  |                       |         |                                      |   |  |            |
| Db            | 7061 GAG 7002  |                       |         |                                      |   |  |            |
| QY            | 682 atgagatgcacaaagggtgtgtgagagaatgtgcctatcagctgaataaacaacatlgag 741  |                       |         |                                      |   |  |            |
| Db            | 7001 GATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6942  |                       |         |                                      |   |  |            |
| QY            | 742 aaggaatgagagcagaagaagagcagcagagaggttcacagacaaacaaactaacagggatcttga 801  |                       |         |                                      |   |  |            |
| Db            | 6941 GAG 6883  |                       |         |                                      |   |  |            |
| QY            | 802 attaaagatcaggaataatgagaatltgagaacttgatgaaagcaacaacacattg 861  |                       |         |                                      |   |  |            |
| Db            | 6882 AAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6823  |                       |         |                                      |   |  |            |
| QY            | 862 gaacagagcagcagaagagcccttgccagagagag 900   |                       |         |                                      |   |  |            |

[illegible]







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:31:03 ; Search time 6491.04 Seconds

(without alignments)  
-432.366 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 2885

Sequence: 1 ggaattccctctgtcgaagt.....aaacaaaagtaaaattt 2885

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_ba2:\*  
3: gb\_cm:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vt:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_cm:\*  
22: em\_ov:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vt:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_ba1:\*  
37: em\_ba2:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

| Result<br>No. | Score  | %<br>Match | Query<br>Length | DB<br>ID     | Description         |
|---------------|--------|------------|-----------------|--------------|---------------------|
| 1             | 2885   | 100.0      | 2885            | 40 AF039691  | AF039691 Homo sapi  |
| 2             | 2762   | 95.7       | 3369            | 40 AF132608  | AF132608 Homo sapi  |
| 3             | 2375.4 | 82.3       | 3450            | 12 AF207748  | AF207748 Mus muscu  |
| 4             | 2375.4 | 82.3       | 3531            | 12 AF006602  | AF006602 Mus muscu  |
| 5             | 2181.6 | 75.6       | 2233            | 9 AB011172   | AB011172 Homo sapi  |
| 6             | 953.6  | 33.1       | 3255            | 40 AF132607  | AF132607 Homo sapi  |
| 7             | 953.6  | 33.1       | 8459            | 9 AB006626   | AB006626 Homo sapi  |
| 8             | 668    | 23.2       | 3758            | 11 HSM800965 | AL117455 Homo sapi  |
| 9             | 625.8  | 21.7       | 2871            | 12 AF207749  | AF207749 Mus muscu  |
| 10            | 289    | 10.0       | 120766          | 40 AC004150  | AC004150 Homo sapi  |
| 11            | 222.4  | 7.7        | 4238            | 9 AB018287   | AB018287 Homo sapi  |
| 12            | 217.2  | 7.5        | 55893           | 56 AC023855  | AC023855 Homo sapi  |
| 13            | 214.6  | 7.4        | 54589           | 42 AC014460  | AC014460 Drosophila |
| 14            | 214.6  | 7.4        | 182015          | 55 AC010993  | AC010993 Drosophila |
| 15            | 203.8  | 7.1        | 1385            | 40 AF124924  | AF124924 Homo sapi  |
| 16            | 173.6  | 6.0        | 3648            | 40 AF132609  | AF132609 Homo sapi  |
| 17            | 173.6  | 6.0        | 4078            | 9 AB020708   | AB020708 Homo sapi  |
| 18            | 173.6  | 6.0        | 4099            | 9 HSA011972  | AL137696 Homo sapi  |
| 19            | 173.6  | 6.0        | 4854            | 11 HSM802148 | AF026202 Caenorhab  |
| 20            | 163    | 5.6        | 3488            | 35 AF128433  | AF026202 Caenorhab  |
| 21            | 163    | 5.6        | 3843            | 12 AF006603  | AC008255 Drosophila |
| 22            | 150.4  | 5.2        | 4435            | 4 XLM112     | AF064207 Schizosac  |
| 23            | 135.2  | 4.7        | 32480           | 34 CELC10E2  | U41410 Schizosacch  |
| 24            | 134.6  | 4.7        | 14737           | 43 AC017824  | M10668 Mouse DNA w  |
| 25            | 134.6  | 4.7        | 113060          | 43 AC008255  | Z50070 Caenorhabdi  |
| 26            | 111.6  | 3.9        | 2750            | 8 AF064207   | Z81488 Caenorhabdi  |
| 27            | 111.6  | 3.9        | 40438           | 7 SP041410   | AL023673 Caenorhab  |
| 28            | 90.8   | 3.1        | 451             | 12 MOSIR3E2  | AC022705 Homo sapi  |
| 29            | 90.4   | 3.1        | 38686           | 34 CEFA43G6  | AC023855 Homo sapi  |
| 30            | 90.4   | 3.1        | 52689           | 32 CECS4H10  | AC004466 Homo sapi  |
| 31            | 90.4   | 3.1        | 153835          | 32 CEY6D1    | AC022705 Homo sapi  |
| 32            | 90.4   | 3.1        | 153835          | 55 AC022705  | AC022705 Homo sapi  |
| 33            | 89     | 3.1        | 155893          | 56 AC023855  | AC004466 Homo sapi  |
| 34            | 88     | 3.1        | 122186          | 11 AC004466  | AC022705 Homo sapi  |
| 35            | 87.6   | 3.0        | 155933          | 55 AC022705  | AC002124 Human BAC  |
| 36            | 87.4   | 3.0        | 100923          | 51 AC002124  | Z71297 S.cerevisia  |
| 37            | 87.2   | 3.0        | 2723            | 7 SCYN021W   | AF091216 Homo sapi  |
| 38            | 86.2   | 3.0        | 97348           | 12 AF091216  | U52064 Kapost's sa  |
| 39            | 84.6   | 2.9        | 3489            | 16 KS052064  | AR065852 Sequence   |
| 40            | 84.6   | 2.9        | 32207           | 5 AR065852   | U75698 Kapost's sa  |
| 41            | 84.6   | 2.9        | 137508          | 16 KSU76698  | AC017028 Homo sapi  |
| 42            | 83.4   | 2.9        | 49745           | 43 AC017028  | AF051726 Mus muscu  |
| 43            | 83     | 2.9        | 5719            | 12 MNCARS1   | U93872 Kapost's sa  |
| 44            | 82.8   | 2.9        | 133661          | 16 U93872    | M10296 Mouse DNA w  |
| 45            | 82.6   | 2.9        | 460             | 12 MOSIR3E1  |                     |

## ALIGNMENTS

```
RESULT 1
LOCUS AF039691 2885 bp mRNA PRI 05-MAY-1999
DEFINITION Homo sapiens antigen NY-CO-9 (NY-CO-9) mRNA, partial cds.
ACCESSION AF039691
VERSION AF039691.1 GI:3170181
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2885)
AUTHORS Gordon,J.D., Tureci,O., Sahin,U., Pfreundschuh,M. and Old,L.J.
TITLE Characterization of human colon cancer antigens recognized by
autoantibodies
JOURNAL Int. J. Cancer 76 (5), 652-658 (1998)
MEDLINE 98272252
REFERENCE 2 (bases 1 to 2885)
AUTHORS Scanlan,M.J.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York
Avenue, New York, NY 10021, USA
FEATURES
source 1..2885
location/Qualifiers
1..2885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/lssue_type="colorectal carcinoma"
<1..2885
/gene="NY-CO-9"
<1..2695
/gene="NY-CO-9"
/note="putative Type II integral membrane protein"
/codon_start=2
/product="antigen NY-CO-9"
/protein_id="AAC18040.1"
/db_xref="GI:3170182"
/translacion="EFLSKSEPTPGSLNHSILPQHPKCGAHNSLIDSSPSPGSP
GTPSYRKLPRGYDSRDPLPKRTASEPLUKRSRKQVQAEERSSPLRRKDYTI
STRKRAVEITGAGPGASVCNSAPGSPSSPNSHTIAENGEGSVPIPIEMHQ
HRAIPDSSNPSLYTSPSLPNTSLGLQATVYTNHSLAPSLSTQDAERQALDS
LRGGTLTGKFMSTSSIPGCLLVALEGDGSPVHNSLQHVLLLEQARQSTLIAP
LHGOSPLVTEGERTVSMRTVGLKPRHNSLRTOSPLRPSQALQDLVMOOQOFLF
KOROOLOLCKITLKTGELPPOPTHEPEHEEELTBOQVLLBGLATLMRSTSES
STQDLEDEBEDEBEDEBEDEDCIOVKDEEGSGABEEDLEPGAGTKLPSDAQPLDP
LOYQVAPLSLATVPHQALGRTQSSPAAPGAKNPDPVPHLETTTSVYDTEMLKHC
LCGNTVHPHAGRIOSIMSRLOETGLSKCERIRGRKATLDEIQTVHSEYHMLKHC
SPLNROKLDSKLLGIPSOQMYAVLPCGGIGVSDVYWNMHSASAVRAVAGCLLEIA
FPAVAGELKNGFALIRPGHAEESTMGCFNSVAITAKLQOKLVNKGVLIVMD
IHNGNCOAFYNDPSVLYLSLRHYDNGPNSGAEFGVSGVGVYNNVMTAGV
DPPIGVEYITARTVYMTIAHESPPVYVSGFDVBEHLSPLEGISTATACFGIL
TROLMTLAGRVVLALEGHDLTAICDASACVSAVLSVLQPLDEGVITQOKPINAV
ATLEKYEIISKMSVCQKFAAGLGRSLRAQAGEEBAEM"
```

BASE COUNT 626 a 901 c 838 g 520 t

Query Match 100.0%; Score 2885; DB 40; Length 2885;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ggaattctctgtcgaaagtaagaagccacacccagcgagcctcaaccattccctcc 60
Db 1 GGAATTCCTCTGTGCGAAGTCAAAGAGCCCAACACGAGCGGCTCAACCATTCCTCC 60
QY 61 acaacacccccaaatgcgcgcgcgcacacatgcctctcttgaccagagttccctccca 120
Db 61 ACAGCACCCCAAATGCTGGGAGGCCACCATGCTTCTTGACACAGAGTTCCCTCCCA 120
QY 121 gagcgccccctggagagcctccctctacaacatgcctcttgctggccctacgacag 180
Db 121 gggcgccccctggagagcctccctctacaacatgcctcttgctggccctacgacag 180
```

```
Db 121 GAGCGCCCCCCTGGGAGCGCTCCCTCTACAAATGCTGCTTGGCTGGGCCCTACAGACAG 180
QY 181 tcgaagcagacttcccccctcgcgcaaaacagcctctgaaccacacttgaagtgcgttaag 240
Db 181 TCGAGACAGACTTCCCCCTCCGCAAAACAGCTCTGAAACCACTTGAAAGTGCCTTCAAG 240
QY 241 gctaaaacagaaggctgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 300
Db 241 GCTAAACAGAAAGGCGGTGAGCGGAGAGAGAGTCCCTCTCGGTGCGCAAGATGGGAC 300
QY 301 tgtattatgaacactttaaagaagaagctgttgaatacaaggctgcgcgcgcgcgcgcgc 360
Db 301 TGTATTATGACACTTTAAGAAGAGAGCTGTGAAATCATCAGATGCGGCGCTGGGCGCTC 360
QY 361 gtcggtggtatcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 420
Db 361 GTCGCTGTATACAGCGGAGCCGCGCTCGGCGCCAGCTCTCCAAACACTCCACAGAC 420
QY 421 catcgctagaatgcttacttgctcagctcagctcccaacaatcccaatgatactccctca 480
Db 421 CATCGCTAGAATGCTTACTGGCTCAGTCCCAACATCCCACTGATGATGCTCCCTCA 480
QY 481 gaaccgagcctccctcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 540
Db 481 GCACCGAGCCCTCCCTCTGAGACAGCTCCCAACAGTTCAAGCTCTACAGCTCTCTTC 540
QY 541 tctgcgcaacatctccctcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 600
Db 541 TCTGCGCAACATCTCCCTAGGAGCTGACAGGCCAGGCTCACTTCAACACTACACCTTAC 600
QY 601 tgcctcccccgaagctgtgcgaacagcagcagcagcagcagcagcagcagcagcagcagc 660
Db 601 TGCCCTCCCGAAGCTGTCGACACAGAGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 660
QY 661 gcaaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 720
Db 661 GCAGGCTGCGACCGCTGACCGCGCAAGTTCATGACACATCTTATCTGCTGCTCTGCT 720
QY 721 gggcgctgcaactgtagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
Db 721 GGGCGTGGCACTGAGAGGCGCGAGGAGCGCCCGACGAGGATCTCTCTGCTGCGACATGT 780
QY 781 gctgttctgtagaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 781 GCTGTGTTGAGAGAGCGCGCGGAGAGAGACACCTTATCTGCTGCTGCTGCTGCTGCTG 840
QY 841 gtccccactagtagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 841 GTCCCACTAGTAGCGGCTGAAGCTGTGGCCACGACATGCGGAGCGGTAGGCAAGCTCC 900
QY 901 ggcgcacatgcgcgccttgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
Db 901 GCGGCATGCGCCCTTGACCGCACTCAGTCTCACCGTGGCGAGATCCCCAGGCGCT 960
QY 961 gcaagcagctgtgtcgtcaacacacacacacacacacacacacacacacacacacacac 1020
Db 961 GCAGCAGCTGTGATGCAACAAACAGACACACAGATCTCTGAGAAGCAGAGCAGCAGCA 1020
QY 1021 gctacagctggggaagatctcctcaagaagaagagagctgcgcgcgcgcgcgcgcgcgc 1080
Db 1021 GCTACAGCTGGGGAAGATCTCTCACCAAGACAGGGAGCTGGCCAGGACGCCACACCCA 1080
QY 1081 cccctgagagacagagagagctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
Db 1081 CCGTGAAGAGACAGAGAGAGCTGACGAGACACAGAGAGTCTTCTGAGGGAGGAGNC 1140
QY 1141 ccttaccatgccccgggagagctcctcaacagagagtgagagcacaaggaagaccttggaga 1200
Db 1141 CCTGACCATGCCCCGGGAGGCTCTCACAGAGAGTGAAGCACAAGGAAGACCTGGAGGA 1200
QY 1201 gggagacgaaggaagagatgagggaaggaaggaagatgtcaatcgaagttaaagaagaga 1260
Db 1201 GGAGAGAGAGAGAGATGGGAGGAGGAGGAGGAGATTCATCCAGGTAAAGCAGAGGA 1260
```

OY 1261 gggcgagagtggtgtctgaaggaagggcccgacttgaggaagccttggtgtctgtgatacaaaaa 1320  
|||||  
DB 1261 gggcgagagtggtgtctgaaggaagggcccgacttgaggaagccttggtgtctgtgatacaaaaa 1320  
OY 1321 actgtctcaaga tgcaccaacgctgcacaactcttgagtgta tccaagcgcccccaacct 1380  
|||||  
DB 1321 ACTGTCTCAGATGCGCAACCGCTGCAACTTWTGCGAGGTGTACAAAGCGCCCTCAGCCCT 1380  
OY 1381 ggcacgtgtgcccaacaaagccctggcgctaaccaatctctccctgtctgccccctggggg 1440  
1381 ggcacgtgtgcccaacaaagccctggcgctaaccaatctctccctgtctgccccctggggg 1440  
DB 1381 GGCCTACTGTGCCCCAACCAAGCCCTGGCGGTACCCAATCTCCCTGCTGCCCTGGGGG 1440  
OY 1441 catgagagagcccccaagacaaacccgttaagacactcttcacaaagaagtgtgtctacga 1500  
1441 CATGAGAGACCCCCAGACCAACCGGTCAAGCACTCTTACACCAAGTGTGTCTACGA 1500  
DB 1501 caagtcatagtctaaagacacaa tgcattgtgcgggaacacacacgtgcacccctgaagcatgc 1560  
|||||  
DB 1501 CACGTTTCATGCTAAAGACACCAAGTCATGTGGGGAACACACACGTGCACCCCTGAGCATGC 1560  
OY 1561 tggccgagatccagagacatctgtgtcccgctgcagagagacagcctgtctagcaagtgcga 1620  
1561 tggccgagatccagagacatctgtgtcccgctgcagagagacagcctgtctagcaagtgcga 1620  
DB 1561 TGGCCGGAATCCAGACATCTGGTCCCGCTGCAGGAGACAGGCTGTGAGCAAGTGCGA 1620  
OY 1621 gcgagatccagagatccgaaagccagcgtatgata tccagacagtgacatctcgaatacca 1680  
|||||  
DB 1621 GCGGATCCGAGGTGCCAAAGCCACGCTAGATGAGATCCAGACAGTGCACCTGTGAATACCA 1680  
OY 1681 caacctgtctctatgggacacagtcacctcaacggcagagaagctcagaacagaagaatgtgtct 1740  
1681 CACCTGTCTCTATGGGACAGTCCCTCTCAACCGGCAAGAGCTAGACAGCAAGAAGTGTCT 1740  
DB 1681 CACCTGTCTCTATGGGACAGTCCCTCTCAACCGGCAAGAGCTAGACAGCAAGAAGTGTCT 1740  
OY 1741 cgggtcccatcaagccsaagaatgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1800  
1741 cgggtcccatcaagccsaagaatgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1800  
DB 1741 CGGTCCCATCAACCCAGAGAGATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800  
OY 1801 tgaacccgtgtgagatgagatgactcctccagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1860  
1801 TGAACCCGTGTGGAATGAGATGACTCCTCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860  
DB 1801 TGAACCCGTGTGGAATGAGATGACTCCTCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1860  
OY 1861 gctgagagctgacctcaagtggt 1920  
1861 GCTGAGAGCTGGCTTCAAGGT 1920  
DB 1861 GCTGAGAGCTGGCTTCAAGGT 1920  
OY 1921 gcccccagagacacaaagccgaggaatccaagccatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1980  
1921 GCCCCAGAGACACACAGCCGAGGAATCCACAGCATGTGATCTGTCTTCAACTGTGT 1980  
DB 1921 GCCCCAGAGACACACAGCCGAGGAATCCACAGCATGTGATCTGTCTTCAACTGTGT 1980  
OY 1981 agccatcaacgcaaaaaactctcaagcagaagtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgt 2040  
1981 AGCCATCAACCCGAAAACTCTCAACAGAAAGTTGAACGTGGCAAGTCTCTCATGCTGA 2040  
DB 1981 AGCCATCAACCCGAAAACTCTCAACAGAAAGTTGAACGTGGCAAGTCTCTCATGCTGA 2040  
OY 2041 ctgggagacattcaacatgtgcaatgtgacacccagagagcgttctcaaatgaccctctgtgtgt 2100  
2041 CTGGGACATTCACCATGGCAATGCAACCCAGCAGCGGTCTTCAATGACCCCTGTGTGT 2100  
DB 2041 CTGGGACATTCACCATGGCAATGCAACCCAGCAGCGGTCTTCAATGACCCCTGTGTGT 2100  
OY 2101 ctacatctctctgtca tgcgtatgacaacggtgaactcttccaaggtctctggggctctccta 2160  
2101 CTACATCTCTCTGCAATGCGTATGACACGGGAATCTTCTTCCAGGCTCTGGGGCTCTCA 2160  
DB 2101 CTACATCTCTCTGCAATGCGTATGACACGGGAATCTTCTTCCAGGCTCTGGGGCTCTCA 2160  
OY 2161 agaggt 2220  
2161 AGAGGT 2220  
DB 2161 AGAGGT 2220  
OY 2221 ggaaccccccatgtgagacgtgtgagtaaccttaacagccttcagagaacgtgtgtgtgtgtgt 2280  
2221 GGAACCCCCCATGTGAGACGTGTGAGTAACTTACAGCCTTCAAGACAGTGTGTGTGTGTGTGT 2280  
DB 2221 GGAACCCCCCATGTGAGACGTGTGAGTAACTTACAGCCTTCAAGACAGTGTGTGTGTGTGTGT 2280  
OY 2281 tggccagagatctcaacatgt 2340  
2281 TGGCCACAGATCTCAACGT 2340  
DB 2281 TGGCCACAGATCTCAACGT 2340

OY 2341 acatctgtctcctctgt 2400  
|||||  
DB 2341 ACATCTGTCTCTCTGT 2400  
OY 2401 ggaactgt 2460  
|||||  
DB 2401 GCAGT 2460  
OY 2461 ggaacgcatctgt 2520  
2461 GACCGCATCTGT 2520  
DB 2461 GACCGCATCTGT 2520  
OY 2521 gcccttgatgtgagagctgttgcagaaaaagcccaacatcaacagtgagtgcaacgttaaga 2580  
2521 GCCCTTGATGAGAGCTGT 2580  
DB 2521 GCCCTTGATGAGAGCTGT 2580  
OY 2581 gaaagtcatcagatccagaacaaacacacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2640  
2581 GAAAGTCATCGAGATCCAGCAAGCAACACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2640  
DB 2581 GAAAGTCATCGAGATCCAGCAAGCAACACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2640  
OY 2641 gggccggtccctgtcgagggcccaagcaggtgagaccgaagaagccgaaatgtgaagcgc 2700  
2641 GGGCCGCTCTGT 2700  
DB 2641 GGGCCGCTCTGT 2700  
OY 2701 atggcctgt 2760  
2701 ATGGCCTGT 2760  
DB 2701 ATGGCCTGT 2760  
OY 2761 cccagccgagagagagccatgtgagcagagacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2820  
2761 CCCAGCCGCGAGAGAGCCCATGTGAGCAAGAGAGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820  
DB 2761 CCCAGCCGCGAGAGAGCCCATGTGAGCAAGAGAGCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2820  
OY 2821 ccccttggt 2880  
2821 CCTTGTGGCTTCAACCATGT 2880  
DB 2821 CCTTGTGGCTTCAACCATGT 2880  
OY 2881 aattt 2885  
2881 AATTT 2885  
DB 2881 AATTT 2885

RESULT 2  
AF132608  
LOCUS AF132608 3369 bp mRNA PRI 06-MAY-1999  
DEFINITION Homo sapiens histone deacetylase 5 mRNA, complete cds.  
ACCESSION AF132608  
VERSION AF132608.1 GI:4754908  
KEYWORDS  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 3369)  
Grozinger C.M., Hassig,C.A. and Schreiber,S.L.  
Three proteins define a class of human histone deacetylases related  
to yeast Hda1p  
Proc. Natl. Acad. Sci. U.S.A. 96 (9), 4868-4873 (1999)  
JOURNAL  
MEDLINE  
REFERENCE  
2 (bases 1 to 3369)  
Grozinger C.M., Hassig,C.A. and Schreiber,S.L.  
Direct Submission  
Submitted (03-MAR-1999) Chemistry and Chemical Biology, Harvard  
University, 12 Oxford Street, Cambridge, MA 02138, USA  
FEATURES  
source  
1..3369  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Jurkat T cell"  
1..3369  
/note="HDAC5; similar to Saccharomyces cerevisiae Hda1p"  
/codon\_start=1  
/product="histone deacetylase 5"  
/protein\_id="AAD29047.1"







|||||  
Db 2304 CCGCCCCCAGCAGCAGAGATGATGCTGTCGCTTGCGGGGGCATCGGGGTGGACAG 2363  
1801 tgaacacgtgtggaatagatgacatccctccagtgctgtgcagatggagtggtcct 1860  
2364 TGACACCGGTGTGGAATAGATGATGATCTCCATGCTGTGCGCATGGCAGTGGGCTCT 2423  
1861 gctgagagctgagcttaagatgtgtcagagagctcaagaatgtttgcatctccg 1920  
2424 GCTGGAGCTGGCTTCAAGGTGGCTGAGGAGGCTCAAGAAAGATTTGCCATCATCCG 2483  
1921 gcccccagagacacacagcccgaggaatccacagcatggatcttgccttccacatgt 1980  
2484 GCCCCAGGACACACAGCGGAGGATTCACAGCATGGATTTGCTTCTTCAACTCTGT 2543  
1981 agccatccacgcaaaaatccctacagagaagtgaacgtggagcaagtgctcatgtga 2040  
2544 AGCCATCACCGCAAAACTCTTACAGCAGAAAGTTGAACGTGGCAGAGGTCCTCATCGTGA 2603  
2041 ctggagcattaccataggaatgcaacagcagcggtctcaaatagaccctctgtct 2100  
2604 CTGGGACATTTCACCATGGCAATGGCACCCAGCGGCTTATTAATGACCCCTCTGTGCT 2663  
2101 ctacatctctctgcatcgctatgacacaggaatctcttccaggtctgtgggtcctga 2160  
2664 CTACATCTCTCTGTCATCGCTATATACACCGGAGACTTCTTCCAGGCTCTGGGCTCTGA 2723  
2161 agaagttgtgtgagagacagcggtgggtgtacaaagtgtgacatgtgacagagagtg 2220  
2724 AGAGTGTGTGTGAGAGACAGCGGTGGGTGTACATGTGAAGTGTGAGCATGAGCAGAGGTGT 2783  
2221 ggaaccccccatgtgagagagtgagtaacttaagaccttaagacagtggtgtatgccat 2280  
2784 GGAACCCCCCATGTGAGAGAGTGTGAGTACCTTACAGCCCTTCAAGCAGTGTGTGATGCCAT 2843  
2281 tggccagaggtctcaacatgagtggttctcagctccgcccgggtttgatgtcttgaag 2340  
2844 TGCCCGAGAGTTCTCACTGATGTGGTCTCTAGTCTCCGCGGTTTGTATGCTTTGAAG 2903  
2341 acatctgtctctctgtgtgtgtactctgtcaccgacagatgttttggccacttgaacag 2400  
2904 ACATCTGTCTCTCTGTGGTGTGTACTCTGTACCGGCAAGATGTTTGGCCACTTGAACAG 2963  
2401 ggaagctgtgagacccctgtgagggggccgggtgtgtgtgtgtgtgtgtgtgtgtgtgt 2460  
2964 GGAAGCTGATGACCCCTGGCAGGGGGCCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3023  
2461 ggaacgcatctgtgagctctgtgaagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2520  
3024 GACCGCATCTGTGATGTGCTCTGAGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3083  
2521 ggccttgtga tgaagcagctctgtgacaaaagcccaacatcaacagcagctgtgacacgtlaga 2580  
3084 GCCCTTGATGAGGAGAGCTGTGACGCAAAAGCCCAACATCAAGCAGTGGCCACGGTAGA 3143  
2581 gaaatcatcagagatccagaaagcaacacatgtgagagctgtgtgtgtgtgtgtgtgtgt 2640  
3144 GAAAGTCATCGAATCCAAAGCAAGCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3203  
2641 ggaacgcatctgtgagggggcccaagcaggtgagacgaggaagcagga -aatgtgaagc 2699  
3204 GGGCGGTGCTGTGAGAGAGCCCAAGAGGTGAGCCGAGAGAGGAGGAGGAGGAGGAGGAGGAGG 3263  
2700 catgtgctgt 2759  
3264 CATGGCTTGT 3323  
2760 ccccaagccgagagagagagcccatgtgagcagagagagctgtgtgtgtgtgtgtgtgtgt 2805  
3324 CCCAGGCGGAGAGAGAGCCCATGTGAGCAGAGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3369

AF207748  
LOCUS AF207748 3450 bp mRNA ROD 07-FEB-2000  
DEFINITION Mus musculus histone deacetylase 5 (HDAC5) mRNA, complete cds.  
ACCESSION AF207748  
VERSION AF207748.1 GI:6911181  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Kao,H.Y., Domes,M., Ordentlich,P. and Evans,R.M.  
TITLE Isolation of a novel histone deacetylase reveals that class I and  
class II deacetylases promote SMRT-mediated repression  
JOURNAL Genes Dev. 14 (1), 55-66 (2000)  
MEDLINE 20107033  
REFERENCE 2 (bases 1 to 3450)  
AUTHORS Kao,H.Y., Domes,M., Ordentlich,P. and Evans,R.M.  
TITLE Direct Submission  
JOURNAL Submitted (22-NOV-1999) Gene Expression Laboratory, The Salk  
Institute for Biological Studies, 10010 N. Torrey Pines Rd., La  
Jolla, CA 92037, USA  
FEATURES  
source  
1..3450  
/organism="Mus musculus"  
/strain="B6:C57BL/6"  
/db\_xref="taxon:10090"  
/dev\_stage="day 14.5 embryo"  
1..3450  
/gene="HDAC5"  
1..3345  
/gene="HDAC5"  
/function="repress transcription"  
/note="class II histone deacetylase"  
/codon\_start=1  
/product="histone deacetylase 5"  
/protein\_id="AA031418.1"  
/db\_xref="GI:6911182"  
/translation="MNSPNESADGMSGRPSLEILPRTPLSHIPVAVEKPYLPGM  
SSMGGGGSPSPVELRGLADPMDPALDEQLOGLVLAQKQOOLQKLOLAEFQK  
HDLTRQHEVQKHLKQOQEMIAARQOELRQORREOORLEKMRKLBQOELLR  
NKEKSESAIATSEVYRLQFLRLSKSRKPTPGGNGSHSPKMGANHASLDOSP  
POSGPPTPSKYRLPLGDPYDRDPLRLKTSAPRLKYRSLKQVARRSSPLLR  
KQIVATSTFKRAVELTGTGPGVSSVCNAPSGSPSSSHSTLAENFTSGVPIT  
TEMTIQHRLPLDSSPNQSLVTSPLSLPAISGLATVTVNTHSLASPKLSTQDAE  
ROALSLRQGTGTGKFMSTSIIPGLLVALEGDSPVGHASLLOHVCSTWRQOST  
LIAVPLHGSPLVTGERVATSMRTVKLIRHNPILSTOSSPLPOSOALQOOLVMOOH  
OOFLKQKQOQOOLKILTKTGLSLRQPTTHPEETEEELIIOQELALIGRGLIIPREG  
SPSESTODLDEEEEEEEDDCIQVDEDESGSPDGPDLSESSAGYKRLPADQ  
QIQLDVQAPLALVPRQALGRTOSSPAAGSMKSPDOPTVVKHLITGVYDTE  
MLKHQCMGNTIHHVHPHAGRIQSIIMSRLOETGLGCEIRGRKATLDEIQVSHYH  
TLUYGSPPLNRKLDKSKLLGPIISQKMYMLPCGGIGVDSIVMNMHSSSAVMAVG  
CLVELAFKVAAGELKNGFALIRPGHHAESTAMGCFNSVAITAKLLOOKLSYGV  
LIVDMQIHNGNTQOAFYNDPSVIYLSLHRYDNGNPPSGAPDEYGGGPGVGNVY  
AMTGCDPRIGVEYITFARIVYMPRLAOFSPDYVIVSGPRVAVBHSPLIGYSTTA  
RCFGHILRLMTLAGRVVLALEGDLTAIDASACYSALLSELQDLDAVLQOK  
PSVNAVATLEKVIETIOSKHMSCVORFPAAGLQSLRDAOTGEKEAEETYSAMALLSVGA  
EOQAVATQHEHSPRAEEMDEPAL"

BASE COUNT 782 a 1031 c 1021 g 616 t  
ORIGIN

Query Match 82.3%; Score 2375.4; DB 12; Length 3450;  
Best Local Similarity 89.8%; Pred. No. 0;  
Matches 2597; Conservative 0; Mismatches 281; Indels 13; Gaps 4;

QY 1 ggaatcctctgtgtgaaatcaagagagccacacagcgccctcaacatccctcc 60  
Db 540 GGAATTCCTGTTTCCAAAGCAAGGAGCCACGAGGGGCTCAACATTCCTCC 599  
QY 61 acagaccccccaaatgtctgtggagagccacacatgtctcttggacacagttccctccca 120  
Db 600 ACAAGACCCCAATGCTGGGAGAGCCACACGCTTCTTTGGACAGAGTTCCCTCCCA 659

|    |      |  |      |
|----|------|--|------|
| OY | 121  | gagcgceccctvgagcgcctccctctcaaaatgctctgctgvgcctcagcag           | 180  |
| Db | 660  | GAGGGCCCTCTTGAGAGCGCTCCCTCTCAAAATGCTTGGTGGCCCTATGACAG          | 719  |
| OY | 181  | tcgagagactccctcccgaaaacagctctgaaacccacttgaagtgcgttaag          | 240  |
| Db | 720  | CCGTATACCTTCCCTCCGTAAACAGGCTCTGGAAACCACTTAAAGTACGTTTCAG        | 779  |
| OY | 241  | gctaaacaagaagtggtctgacgaggaacacagctccctccctgcgtgcgaagaatgagac  | 300  |
| Db | 780  | GCTTAAACAGAGGTACCAGAGAGAAACATCCCTCTCTGCTGCAAGAGATGGAC          | 839  |
| OY | 301  | tgtattagcaacttaagaagagctgtltgagatcacaggtgcggcgctgvggcgc        | 360  |
| Db | 840  | TGTTATTAGTCTTTAAAGAAAGACAGTTGAGATCACGGCACGGGGCTGGGGGTGTC       | 899  |
| OY | 361  | gtccggtgtgtaacgacgaccggtctccgccccactctcccaacagctcccaagaac      | 420  |
| Db | 900  | GTCGTGTGTAAAGTCCGCGCGGCTCTGGCCCCACCTCTCCCAACAGTCCACAGAC        | 959  |
| OY | 421  | catcgctgagatgtctctactgctcagtlccccaacatcccaactgagatgtctccca     | 480  |
| Db | 960  | CATGCTGTAAAGACGCTTACTGTGGCTACGTCCCAACATCCCACTGAGATGATCCCCA     | 1019 |
| OY | 481  | gaaccgagccctccctctgacagctcccccacacagttcaagccttaacagctcttc      | 540  |
| Db | 1020 | GCACCGGGCCCTCCCTCTGGACAGATTCCCAACCGTAGTCACTGTATAGCTCTTC        | 1079 |
| OY | 541  | ctctgccaacatctcccttaagctgtcagggcgaagctcatctgtcaaccaactcaactac  | 600  |
| Db | 1080 | TCCTGCCAACATCTCCCTAGGGCTGACAGGCACTGTACATGTACCAATCCCACTCAC      | 1139 |
| OY | 601  | tgccctcccgaaagtgtgcagacagcaagagggcgagagagagccctcagctccctcg     | 660  |
| Db | 1140 | CGCTCCCGAAGCTGTCAACACAGCAGAGGCTGAGAGGACGCCCCCTTCACTCCCTCG      | 1199 |
| OY | 661  | gcagagtgtgcaagcttgcacggcgaagtcatgtagcaatcctatctctctgctgctct    | 720  |
| Db | 1200 | GCAGGGGGGACACTGTGACCGGCAAGTTCAATGAGACATCTCATCCCTGCTGCTGT       | 1259 |
| OY | 721  | ggcggtgacactgtagagggcgagggagccccacgggcaatgcctccctgctgcaagatgt  | 780  |
| Db | 1260 | GGGAGTGGCACTGTGAGGGGTGACACAAAGCCCCACGGGGACGCTTCCCTGTGCACACGT   | 1319 |
| OY | 781  | gctgtgtgtagaagggccggcagcagagacccctcatctgtgtgtgtgtgtgtgtgtgt    | 840  |
| Db | 1320 | TTTGCTCTGGACA---GGCCGGCAACAGACACGCTCATAGCAGTGGCGCTTCATGGCA     | 1376 |
| OY | 841  | gtccccaactagtgaacgggtgaacgtgtgtgtgccaacagcatgcygaagctgaagctcc  | 900  |
| Db | 1377 | GTCCCCACTGTGTGACGGGTGAACGTGTGGGCACACAGCATGAGGACGGTGGTAAGCTCC   | 1436 |
| OY | 901  | ggcgcatcgggcctctgagccgcaactcagtctcaaccgctgcgcgagaggtcccaaggcct | 960  |
| Db | 1437 | GAGGACACGACCTCTGTGAGCCGACACTAGTCTCTCCCGCTGGCGAGAGTCCCAAGGCTCT  | 1496 |
| OY | 961  | gcagcagcttgtcatgtaaaacagacacagcagcttccctggagaaagcagacagcaca    | 1020 |
| Db | 1497 | GCACACAGTGTGCTGTGAGACAGACACCGACGTTCTCTGGAGAAAGCAAGAACACACCA    | 1556 |
| OY | 1021 | gtcacagctggtgcaagatctctcaaccaagacagggagctgtcccaagcagccacaccca  | 1080 |
| Db | 1557 | GATGCAGCTGGGCAAGTCTTACCAAAACTGGGGAGCTGTCAAGGCAAGCCACACACTCA    | 1616 |
| OY | 1081 | cccttgagagacagagagagagctgacgggacagcagagagctctgtctgvgggagagac   | 1140 |
| Db | 1617 | CCCGAGAGACACAAAGAGAGCTGACGAGCAGCAGAGGCTTGTGTGGAGAGGGGGC        | 1676 |
| OY | 1141 | ccctgacatgcccgggagaggtctccacagagatgtgagacacagaagaactgtaga      | 1200 |
| Db | 1677 | CCCTACCACTTCCCGGGAGAGCTCTACAGAAAGTGTAGACACCCAGAGAAACCTAGAGA    | 1736 |

|    |      |  |       |
|----|------|--|-------|
| QY | 1201 | ggaggaaggaagaagatgaggagaggaaggaatgcaaccaggttaagcagaga                    | 1260  |
| Db | 1737 | ggagagagagagagagaga-----ggaggaagagagagagagctcattccaggtcaagagagaga        | 17933 |
| QY | 1261 | ggcggaagatggtcttaaggaagggcccagctggagagagcttggtctgagataaaaaa              | 13200 |
| Db | 1794 | tggcagagtggtccttatgaaagcccttgaccttgaagagttcgaagttcagttgattacaaaaa        | 18533 |
| QY | 1321 | actgtctccaatgtcccaaacgcgtctgaaccttggacgtgtgtaaccaagcgcccccaacct          | 13803 |
| Db | 1854 | gwtgttgcacaaatccacagacagatttacagcccttcacaggttgcacagagacccttcagctt        | 19133 |
| QY | 1381 | ggcaactgtgcccacaacaagccttggtgcgttacccaactctccctgtctgccttgagg             | 14400 |
| Db | 1914 | ggccacctgtgctctatcattcagagcccttgaggccgcacccagtcctcactcgctgcttggag        | 19733 |
| QY | 1441 | catgaagaaccccccaacaaaccc-----gtcaagacaccttcaaccaagaatgtgt                | 14944 |
| Db | 1974 | catgaagaagcccccaacaaacacccactgtgggtgaagcactcttccacacaggtgtgt             | 20333 |
| QY | 1495 | ctacgaacgttctcatgttaagacacacagatgcatgtgtcgggaacacacacgttgacacctga        | 15544 |
| Db | 2034 | ctatgacacacgtttcatgtgtgaagcaccagttatgttcggaaaaacacacgttcacaccaga         | 20933 |
| QY | 1555 | gcatgtctgcccgaatcacagacatctgtctccggtctcagagacagcgcttgatagca              | 16144 |
| Db | 2094 | gcacgcgcgcgcgcattccagagacattgtgtcccgctgcagagaaactgtgtcttcggcaa           | 21533 |
| QY | 1615 | gttgagagcgatctcgaagtgtcgaagaagccagctcagttgatccagaagaatgtacctcga          | 16744 |
| Db | 2154 | gtgtgagcggaattccgggggtgttaagccacactggaatgaattccagaccgttcacttgcga         | 22133 |
| QY | 1675 | ataccaacccctgtctcatgtggaacagctcccccacaacggcagagaagctagaagacaaga          | 17344 |
| Db | 2214 | gtaccacacacccctgcttattatggagacagccccccttaaccggcagaagcttgagacagcaaa       | 22733 |
| QY | 1735 | gtgtctcgtgtccatcagcagcaagaatgatgtctgtctcgtctgttggtggatcctgaggt           | 17944 |
| Db | 2274 | gctctgttgcccccattcagccacaagaatgtgtacgccattgctcccttgaggggcattggcgt        | 23333 |
| QY | 1795 | ggacaatggaacacggtgaggatgagatgtcaatctctcaatgtctgtgagcatgagcaatgtgg        | 18544 |
| Db | 2334 | ggacaatggaacacggtgtggatagatagatgtcaatctctcaatgtctgtgagcatgagcaatgtgg     | 23933 |
| QY | 1855 | ctgctctctgtgaagctgtgcttcaagaatgtgctgagagagagcccaagaatgtgaattgcat         | 19144 |
| Db | 2394 | ctgctctgttgagagctgtgaccttcaaaaggtgctgtagagagagctcmaagatgttgatctcat       | 24533 |
| QY | 1915 | catccggcccccaaggacacacgcgcgaagaaatccacagccatggagatctgcttctcaa            | 19744 |
| Db | 2454 | catccggcccccaaggacacacgcgtggaatagatgtcaatctctcaatgtctgtgagcatgagcaatgtgg | 25333 |
| QY | 1975 | ctctgtgtgcatcaacgcgcaaaactctctaagaagaaatgtgaacgtgggcgaaggtctcat          | 20344 |
| Db | 2514 | cttccgttggacccaatcacaatcttctgtcagcagaacgtcgtgagcagtgagccatgagctctcat     | 25733 |
| QY | 2035 | cgatgactgtggaacatcaacatgagaaatggaacccagcagcggttctcaatgtgacccctc          | 20944 |
| Db | 2574 | cgttgactgtggatatttccacattgacacagcagccacccacagacattcttcaacagatccctc       | 26333 |
| QY | 2095 | tgatgtctatacatctctctgtatcgtctaigtacaacagggaaacttcttccaggtctctgggc        | 21544 |
| Db | 2634 | tgtgtcttacaattccctctgcatgctgtaacgcaaacgggaacttctttccaggtcttggggc         | 26933 |
| QY | 2155 | tctctgaagaagttgtgtgaggaagccagcgttggtgtacaaatgtgaaagtgtgcattgagcaag       | 22144 |
| Db | 2694 | tctctgaagaaggtgtgtgtgaggaagccagcgttgggtgtacaaactgtgaaagtgtgcattgagcaag   | 27533 |
| QY | 2215 | aggtgtgtgaaccccccaattggagaacgttgagaacttaagaacttcagagcttcagagaatgtgat     | 22744 |
| Db | 2754 | aggtgtgtgaatcccccaattgtggaatgtggaatgtcgaacagccttccagagacagtggtgat        | 28133 |
| QY | 2275 | gccatttgccacagatgtcttcaacctatgtgtgtcctagtctccgcgcgggttctgtatgctgt        | 23344 |

|            |  |  |             |
|------------|--|--|-------------|
| Dd         | 2814   | GCCCAATGCCAGGAGTTCTACCGTAGCTGTCTAAGTCTTCGGTGGGTTGAAGTCTGT      | 2873        |
| Oy         | 2335   | tgaagacaatctgtcttcctcttggtygactactctgtcaaccgcagaatgttttgccactt | 2394        |
| Dd         | 2874   | TGAAGACATCTGTCTCCACTGCGTGGGTATTCTCTCACCGGCNAATGTTTGGCCACTT     | 2933        |
| Oy         | 2395   | gaacagcagctgatatacccttgcaggaggccgggtggtgtctgtcccctggaggagcca   | 2454        |
| Dd         | 2934   | GACCAGCAGCTCATAGACACTGGCTGGGGGGCCGGGTGGTGTGGCTCGAGAGAGGCCA     | 2993        |
| Oy         | 2455   | tgacttgaccgcacactgttatgacctcttgaagcttgttctcgctctgcttaagttaa    | 2514        |
| Dd         | 2994   | TGACTTGAACCGGCATCTGTGAAGCCTCTGAGGCCCTGTCTGGCTGTGCTAGCGTGA      | 3053        |
| Oy         | 2515   | gttgtagcccttggatggagcagctcttcagaanaagcccaacagctcgaagctggccac   | 2574        |
| Dd         | 3054   | GCTGCAGCCCTTTGGATGTAAGCAGTCTTTCAGCAAAGCCAGCCTCAATGCGTTGCCAC    | 3113        |
| Oy         | 2575   | gctagaagaatcatcagatattccagaagaacactgtgactgtgtgcagaagltgcgcgc   | 2634        |
| Dd         | 3114   | ACTAGAAGAATCATCGAGATTCAGAGCAAAACACTGAGCTGTGTACAGAGTTTGCCGC     | 3173        |
| Oy         | 2635   | tggtctggccggtccctctgcagaggccccaaagcaggttagaccgaagaagccga-aatgt | 2693        |
| Dd         | 3174   | TGGTTGGGCTGTCTGTCTGCTGGGAGGCTCAGACAGTGTGAAGAAGAGAGGCCGAGACTGT  | 3233        |
| Oy         | 2694   | gaagccatgtgctgtcgt   | 2753        |
| Dd         | 3234   | GAGCCCATATGGCCCTTGCTTTCCTGTGGGGCTGTAGAGAGGCCAGGCTGTGGCCTCAMA   | 3293        |
| Oy         | 2754   | acacagcccccaagccgcagagagaagcccatgtgagcagagacctgtcctgtgaagccgcg | 2813        |
| Dd         | 3294   | GCACAGCCCCCAGCCAGCAGAGAGAGCCCATGTAGCAGAGAGGCTGCTGTACACCTTG     | 3353        |
| Oy         | 2814   | cccccatcccttgggtctaaccatlfgatttittatltttctatataaacaaca         | 2873        |
| Dd         | 3354   | CCCCATCCCTTTGGGCTTCATCATTTGATTTTATTTTTCTATAAAAAACAAA           | 3413        |
| Oy         | 2874   | agttaaaaat 2884  |             |
| Dd         | 3414   | AGTCACACATT 3424   |             |
| RESULT     | 4  |  |             |
| AF006602   |  |  |             |
| LOCUS      | AF006602   | 3531 bp  | mRNA        |
| DEFINITION | Mus musculus histone deacetylase MHD1 mRNA, complete cds.          |  | 15-FEB-2000 |
| ACCESSION  | AF006602   |  |             |
| VERSION    | AF006602.2   | GI:6978313   |             |
| KEYWORDS   | .  |  |             |
| SOURCE     | house mouse.   |  |             |
| ORGANISM   | Mus musculus   |  |             |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;      |  |             |
| AUTHORS    | Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.           |  |             |
| TITLE      | 1 (bases 1 to 3531)  |  |             |
|            | Verdel,A. and Khochbin,S.  |  |             |
|            | Identification of a new family of higher eukaryotic histone        |  |             |
|            | deacetylases. Coordinate expression of differential transcription- |  |             |
|            | dependent chromatin modifiers                                      |  |             |
| JOURNAL    | J. Biol. Chem. 274 (4), 2440-2445 (1999)                           |  |             |
| MEDLINE    | 99107904   |  |             |
| REFERENCE  | 2 (bases 451 to 3531)  |  |             |
| AUTHORS    | Verdel,A. and Khochbin,S.  |  |             |
| TITLE      | Direct Submission  |  |             |
| JOURNAL    | Submitted (02-JUN-1997) Institut Albert Bonniot, INSERM U309, La   |  |             |
| REFERENCE  | Tronche 38706, France  |  |             |
| AUTHORS    | 3 (bases 1 to 3531)  |  |             |
| TITLE      | Verdel,A. and Khochbin,S.  |  |             |
| JOURNAL    | Direct Submission  |  |             |
|            | Submitted (15-FEB-2000) Institut Albert Bonniot, INSERM U309, La   |  |             |
|            | Tronche 38706, France  |  |             |
| REMARK     | Sequence update by submitter                                       |  |             |

[illegible]

QY 481 gacccagacccctccctctgagacagctcccccacacagttccagctctctcaagctctctc 540  
 Db 1101 GCACCGGGCCCTCCTCTGAGACAGTCCCAAAACAGTTCCAGCTCTATAGCTCTCTTC 1160  
 QY 541 tctgcccacatctctctctagggctgagagccaggtctcactgtcccaactcaacctac 600  
 Db 1161 TCTGGCCCAACATCTCTCTTAAAGGCTGACAGCCACTGTCACTGTCCACCAACTCCACTTAC 1220  
 QY 601 tgcctcccagagctgtctgacacagcagagagccagagagccctccagctccctcg 660  
 Db 1221 GCGCTCCCCGAAGCTGTATACACACAGAGAGGCTAGAGAGCCCTTCACTCTCCCTGCG 1280  
 QY 661 gcaaggctgacagctgagccggaagttcatgagcacatctctctctctctgctgct 720  
 Db 1281 GCAGGGCGGCACACTGACCGGCAAGTTCATGACATCTCTCATCTCCCTGGCTGCTGTT 1340  
 QY 721 gggcgtgacatgagagggcgagagagcccccagagcagatgctccctctgctgagcagt 780  
 Db 1341 GGGAGTGGCACTGGAGGGGTGACACAAAGCCCCACGGGACGCTTCCGTGTCACACAGT 1400  
 QY 781 gctgtgtgtagcagagcccgagcagagcacctctctctctctgtgtgacatccacgggca 840  
 Db 1401 TTGCTCTCTGACA--GGCCGGCAACAGACACGCTCTCATAGCAGTGGCTTTCATGGGCA 1457  
 QY 841 gtcccccagtgagcgggtgagcgtgtgagcagcagcaltgagcagcgttgaagcagctcc 900  
 Db 1458 GTCCCACTGGTGGAGCGGTGAACGTGTGGCCACGACATGAGAGAGGTTGGGTAAGCTCCC 1517  
 QY 901 ggcgacatcgcccttgagcgcagcactcagctcctcacgcgtgcgcagagctcccaaggcct 960  
 Db 1518 GAGGACGACGACTCTGACCCGACACACATCACTCTCCCGCTGCGGAGATCCCAAGCCCT 1577  
 QY 961 gcaagcagctgtcatgagcagacacagcacagcagctctctctgagagagcagagcagcagca 1020  
 Db 1578 GCAGCAGCTGGTCTATGACAGCAGCAGCAGCAGCAGTTCCTGGAGAAAGCAAGAAAGCAGCAGCA 1637  
 QY 1021 gctaagcttgggcaagatctcctcaacaaagacaggggagcgtgcccagagcagcagcagcagca 1080  
 Db 1638 GATGAGCTGGGCAAGATCTTACCAAAATGGGGAGCTGTCAAGAGCAGCCACACACCA 1697  
 QY 1081 cccctgagagagacagagagagcgtgacgagagcagcagagagcttctgtctggggagagagc 1140  
 Db 1698 CCGGAGAGAGACAGAAAGAGAGCTGACAGCAGCAGCAGAGAGGCTTGTGGAGAGAGGGGCG 1757  
 QY 1141 cctgacacatgcccgggagggctccacacagagagtgtagagcacacagagagagccttgagagca 1200  
 Db 1758 CCTGACCATTTCCCGGGAAGGCTCTACAGAAAGTAGAGCACCCAGAGACCTTAGAAGA 1817  
 QY 1201 ggaagacagagaaagagagctgggagagagagagagatgcatlccaggttaagagcagagca 1260  
 Db 1818 GGAGGAGAGAGAGAGGA--GGAAGGAAGAGAGAGATGTCATTGACGTCAGAGATGAAGA 1874  
 QY 1261 gggcagagagtgctgtgagagagggccgagcttgagagagcctgtgtctgatacaaaaa 1320  
 Db 1875 TGGGAGAGGTGTCTGTATGAAGGCCCTGACTTAAAGACTTAAAGTGTGTATCAAAAAA 1934  
 QY 1321 actcttccagatgcccacacagcgtgcaaccttggagagtgtagcaagagggccctcagact 1380  
 Db 1935 GTTGTTCCTCAATGCCAGCAGTTTACAGCCCTCCAGGTGTACAGAGACCCCTCAAGCT 1994  
 QY 1381 ggcacatgtgcccacacagcccttggcgttaccatctccctctgtcctctggggg 1440  
 Db 1995 GGCCACTGTGCTCATCAAGGCCCTGGGGCCGACCCAGTCTGCTGCTGGGAG 2054  
 QY 1441 catgagaagaccccccacagcaacc-----gtcaagcactctcaccacaagtgtgt 1494  
 Db 2055 CATTAAGAGCCCAACAGCAGCAACACACATGTGTGAAGCACTTCTTACCAACAGGTGTGT 2114  
 QY 1495 cttagagacgttcatgtctaagcaccagtgcatgtgagggagacacacagctgacccctga 1554  
 Db 2115 CTATGACACGTTATGCTGAAAGCACCAAGTGTATGTGGAAACACACAGTGACCCAGA 2174

QY 1555 goatgtgcccggatccagagcatctggtcccgctgagagagacaggcctgtatgacaa 1614  
 Db 2175 GCACCGCGCCGATCCAGAGCATGTGTCCCGGCTGACAGAAATGGTCTGCGCAA 2234  
 QY 1615 gtgagagcgatccgaggttcgcaaaagccagcgttagatgccagacagtgcaactctga 1674  
 Db 2235 GTGTAGGGGATCCGGGTCTGAAAGCCACACTGGATGAATTCAGACCGGTGCACTCTGA 2294  
 QY 1675 ataccacccctgtctctatgtgagccagctccctcaaccggagagagctagaagcaagaa 1734  
 Db 2295 GTACACACCTGCTCTATGGAGCAGCAGCCCTTAAACGGCAGAAAGCTGAGACAGCAAA 2354  
 QY 1735 gtgtcgtgtcccatcagccagaagatgtagtgtgtcctctgttgggggacatcggtgt 1794  
 Db 2355 GCTGCTTGCCCATCAGCCAGAAAGATGTAGCCATGCTGCGCCTGTGGGGCATGTGGGGT 2414  
 QY 1795 ggaacgtgaacacggtgtgaatgtagtcatctctccagctgtgtggtgcataggcaatggg 1854  
 Db 2415 GGACGTGACAGGGGTGTGAATGAGATGCACTCTCAAGTSCCGTGCATATGGCAGTGGG 2474  
 QY 1855 ctgcctgtgagctgagccttcaaggtgtgctgcaagagagctcaagaatgagattgacct 1914  
 Db 2475 CTGCTGTGTGAGAGCTGGCTTCAAGGTGGCTGACAGAGAGCTCAAGATGGATTGTCTAT 2534  
 QY 1915 catccggcccccagagacacacacgcccaggaatcccaagcatalgagattctctctcaa 1974  
 Db 2535 CATCCGGCCCCAGGACACCATATGTGAGAGTCCACACCATATGGGATCTGCTTCAA 2594  
 QY 1975 ctctgtagccatccagcgaaaatctctacagcagaagttgaacgttgggcaagttccat 2034  
 Db 2595 CTCCGTACCATCAACAGGTAACTCTGACACAGAACTGAGCGGTGAGGCAAGTCTCTAT 2654  
 QY 2035 cgtgagcctggagacattcaacatggaatggaatggcaccagcagcggttctcaatgaaacctc 2094  
 Db 2655 CGTGACCTGGGATATTCACCATGGAAGGACACCCAGACATCTTCAACAGATCCCTC 2714  
 QY 2095 tgtgtctacatctctctctgcatctgcatlgaacacgggaactcttccagcctctggg 2154  
 Db 2715 TGTGCTTACATCTCCCTGCAATCGCTACGACAGGAGAACTCTTCCAGGCTCTGGGGC 2774  
 QY 2155 tccctgaagaagttgtgtgagagacacagcggtgtgtatcaaatgtgaactgtgcatggaagc 2214  
 Db 2775 TCTGTAAAGAGTGTGTGAGGAGGAGGAGGTGGGTGATCAACCTTAATGTGGGTGACAGG 2834  
 QY 2215 aggtgtgagcccccacatttggagagcgttgagtaaccttcaagccttcaagagcagtgat 2274  
 Db 2835 AGGTGTGATCCCCCATTTGGAGATGTGAATACCTGACACACTTCAAGAGAGTGTGAT 2894  
 QY 2275 gcccatgtcccagagttctcactgcatgtgtgtcctagctccgcggtttagatgt 2334  
 Db 2895 GCCCATTTGCCAGGAGTCTCACCTGACGTGCTCTACTCTCCGCTGGGTTGATGTGT 2954  
 QY 2335 tgaaggaatctgtctctctcgtgtgtgtactctgtcaacgcccagatgttttggccaact 2394  
 Db 2955 TGAAGGACATCTGTCTCCACTGTGGGTGCTATTTGTCTCCGCGAGATTTTGGCCACTT 3014  
 QY 2395 gaccagagcagctgtagacccctggcagggcggggtgtgtgtgttggccctggagggagagcca 2454  
 Db 3015 GACCAAGAGCTCATGACACACTGGCTGGGGGCGGGTGTGCTGGCCCTGGAGAGAGGCA 3074  
 QY 2455 tgaactgacgcatctgtgtgtctctgtgaagctgtgtctcgtgctgtcagtgtaaa 2514  
 Db 3075 TGACTTGAACCCCATCTGTGTATGCTCTGAGAGCCTGTGTCTCGGCTGTGTCACAGTGA 3134  
 QY 2515 gtctgagcccttggatgagggagctcttgagcaaaagcccaacatcaacgagatggccac 2574  
 Db 3135 GCTCAGGCCCTTGGATGAAGAGTCTTGACAAAGCCCAAGCCGCTCATATGGGTTGGCCAC 3194  
 QY 2575 gctagagaagatcaatcgagatccagagcaaacacttgagcgtgtgtgcaagttcggccg 2634  
 Db 3195 ACTAGAGAAATCATCTGAGATCCAGAGCAAAACACTGAGAGCTGTGTACAGAGTTTGGCCG 3254  
 QY 2635 tgtctgtggccgtctcctgtgagggggcccaagcaggtgagagccgaagaagcgca -aatgt 2693

| DB         | Accession  | Version    | Keywords | Source | Organism    | Reference Authors Title | Journal Medline Features | Gene | CDS |
|------------|------------|------------|----------|--------|-------------|-------------------------|--------------------------|------|-----|
| Db         | 3255       |            |          |        |             |                         |                          |      |     |
| Oy         | 2694       |            |          |        |             |                         |                          |      |     |
| Db         | 3315       |            |          |        |             |                         |                          |      |     |
| Oy         | 2754       |            |          |        |             |                         |                          |      |     |
| Db         | 3375       |            |          |        |             |                         |                          |      |     |
| Oy         | 2814       |            |          |        |             |                         |                          |      |     |
| Db         | 3435       |            |          |        |             |                         |                          |      |     |
| Oy         | 2874       |            |          |        |             |                         |                          |      |     |
| Db         | 3495       |            |          |        |             |                         |                          |      |     |
| RESULT     | 5          |            |          |        |             |                         |                          |      |     |
| LOCUS      | AB011172   | 2233 bp    | mRNA     | PR1    | 10-APR-1998 |                         |                          |      |     |
| DEFINITION | AB011172   |            |          |        |             |                         |                          |      |     |
| ACCESSION  | AB011172.1 | GI:3043723 |          |        |             |                         |                          |      |     |
| VERSION    | KIAA0600   | protein.   |          |        |             |                         |                          |      |     |
| KEYWORDS   |            |            |          |        |             |                         |                          |      |     |
| SOURCE     |            |            |          |        |             |                         |                          |      |     |
| ORGANISM   |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          |        |             |                         |                          |      |     |
| AUTHORS    |            |            |          |        |             |                         |                          |      |     |
| TITLE      |            |            |          |        |             |                         |                          |      |     |
| JOURNAL    |            |            |          |        |             |                         |                          |      |     |
| REFERENCE  |            |            |          | </     |             |                         |                          |      |     |

|   |              |   |            |             |        |        |
|---|--------------|---|------------|-------------|--------|--------|
| LCGTSVYARCGCHITROLMTLAGRVVIALEGGDLVRIADASACYSALLSVBLDTL<br>DEAAVQDPNINAVATLEKVEITQISKMSVCQFKAAGUGNRLRQAQGEATEATVSA<br>MALLSVGEQAOAAARSHRPAEPEEMEDPAL" | BASE COUNT   | 469 a   | 663 c      | 702 g       | 399 t  | ORIGIN |
| Query Match   | 75.6%        | Score 2181.6  | DB 9       | Length 2233 |        |        |
| Best Local Similarity   | 98.9%        | Pred. No. 0   |            |             |        |        |
| Matches 2207  | Conservative | 0   | Mismatches | 24          | Indels | 1      |
|   |              |   |            | Gaps        | 1      |        |
| QY  | 654          | cccttcgagcaagggtgtggcaagcttgaccggcaagtttaataagcaatcccttatctctgtgc | 1133       |             |        |        |
| Db  | 1            | CCCTCGGAGGAGGTGGACACGTGACCGGCAATTATGAGCAATCCTTATCTCTGCT           | 60         |             |        |        |
| QY  | 714          | gacctctggcgttggcacttgagagggcgacggagagcccccacggagatgctctccctgcgc   | 773        |             |        |        |
| Db  | 61           | GCCTCTGGGGCTGGCAGCTGGAGGGCGACGGGAGGCCCCACCGGGATGCTCTCCCTGCTGC     | 120        |             |        |        |
| QY  | 774          | agcattgtcgtgtgtcttgtagagcagggcccgagcaagagcaacctcaattgtcgtgcactcc  | 833        |             |        |        |
| Db  | 121          | AGCATGTGCTGTGTGCTGGAGAGCGCCCGGACAGAGACACCTCATGTGCTGGCCACTTC       | 180        |             |        |        |
| QY  | 834          | aaggcgagttcccaactatgtgaaggggtgaacgtgtgtgccaccagcaatcggaacgttaagca | 893        |             |        |        |
| Db  | 181          | AGGGGAGTCCCACTAGTATGAGCGGGTGAACGTGTGGCCACAGCATGTGGAGGATAGCA       | 240        |             |        |        |
| QY  | 894          | agctcccgagagatatggcccccttgagcccgcaactcagttctcaaccgtctgcgaagatccc  | 953        |             |        |        |
| Db  | 241          | AGCTCCCGGGGATGGCCCCCTTGAGCCGCACTCAATCTTCACCGCTGCGCGAGATCCCC       | 300        |             |        |        |
| QY  | 954          | aggccctgcagagcattgtgtcatgtacaaacaagacacagcaatctccttgtagaagcaagc   | 1013       |             |        |        |
| Db  | 301          | AGGCGCTGGAGCAGCTGGTGCATGCAACAACAGACACAGCATGTTCTGGAGAAAGCAAGC      | 360        |             |        |        |
| QY  | 1014         | agcagcagctacagcttgggcaagatctcaccagaacagggagagctgcccaggaagcca      | 1073       |             |        |        |
| Db  | 361          | AGCAGCAGCTACAGCTGGGGCAAGATCCTCACCMAGACAGGGAGCTGCCCAAGGCCCA        | 420        |             |        |        |
| QY  | 1074         | ccaccaccccttagaagagcaagagggaggtctgagagcagaggaaggttcctgtcgaggg     | 1133       |             |        |        |
| Db  | 421          | CCACCCACCTCTAGGAGAGACAGAGAGAGCTGACGAGCAGCAGGAGGCTTGTGCTGGGG       | 480        |             |        |        |
| QY  | 1134         | aaggagcccttgacaaatgcccgcggagggctccacagagagttaagagcaacaagaaagcc    | 1193       |             |        |        |
| Db  | 481          | AGGGAGCCCTGAGCATGGCCCCGGGAGGGCTCCACAGAGAGTCAAGACACAGAGAGAC        | 540        |             |        |        |
| QY  | 1194         | tggagagagagagacgaagaaagagatctggagagaggaagatgtcatccaagttlaag       | 1253       |             |        |        |
| Db  | 541          | TGGAGGAGAGAGACAGACAGACAGATGGCGAGAGAGAGAGATTCATCCAGCTTAAAG         | 600        |             |        |        |
| QY  | 1254         | acgagagagggcgagaggtgtgtcttgaggaggggccgacttggagagagctgtgttgat      | 1313       |             |        |        |
| Db  | 601          | ACGAGGAGGGCCAGAGTGTGTGCTGAGGAGGGGCCCGACTTGGAGGAGCCTGGTGTGAT       | 660        |             |        |        |
| QY  | 1314         | acaaaaaaactgtcttcacagatgcccacacgctbcaacctttgaaagtgttacaagaagcccc  | 1373       |             |        |        |
| Db  | 661          | ACAAAAAATGTTCTTCAGATATGCCAGGCCGCTTCAGACCTTTGGAGGTATACAGGGGCC      | 720        |             |        |        |
| QY  | 1374         | tcagctctggcactgttgcccacaacaaagcccttggcgttaccatccctccctgtgcc       | 1433       |             |        |        |
| Db  | 721          | TCAGGCTGGCCACTGTGGCCCCACAGGGCCCTTGGCGCTTACCAAGTCTCCCTCGTGGCC      | 780        |             |        |        |
| QY  | 1434         | ctgggggagcatgaagaaccccccgagaccaaccgtaagaaccttttaacaaagtgtgg       | 1493       |             |        |        |
| Db  | 781          | CTGGGGGCGCTGAAGAGCCCCCCAGACCAACCCGCTTAAGACCTTTTACACACAGAGTGG      | 840        |             |        |        |
| QY  | 1494         | tctacgaacgcttcaatgttaagaacacagttgatgtcggaagaacaacagtgacccgt       | 1553       |             |        |        |
| Db  | 841          | TCTTAGACACGTTCACTGCTTAAAGACACAGTGCATGTGGCGGAACACACAGTGCACCTG      | 900        |             |        |        |
| QY  | 1554         | acgatgtctggccgagatcccaagacatctgtgtcccggtcgagagagaacagctgttaaga    | 1613       |             |        |        |

|   |   |      |  |      |
|---|---|------|--|------|
| D | b | 901  | AGAMTGTGGCCGGATTCAGAGCATCTGTGTCGCCGCTGCAGAGAGACAGGCTCGTTAGCA         | 960  |
| O | y | 1614 | atgctgagcggatcccgagctcgcaaaagccacgcttagatgagatccagacagtgcactctg      | 1673 |
| D | b | 961  | AGTGGAGCCGGATCCGAGGTCTGCAGAAAGCCACGGCTAATGTAGATCCAGACAGTGCACCTCG     | 1020 |
| O | y | 1674 | aatccacaaccctctctatgggagccagctccctcaaccggcaagatcaagcacaaga           | 1733 |
| D | b | 1021 | AATACCAACCCCTCTATTATGGGACCAAGTCCCTCAACCGGCGAAGCTATGACACACAAG         | 1080 |
| O | y | 1734 | agttgctcgctcccatcaagcacaagaatgatctgtgctcccttgctgtgttgaggacacggg      | 1793 |
| D | b | 1081 | AGTTGCTCGCGCCCATATGACCAAGAAATGATATCTGTGCTGCTTGTGTGGGGCATCCGGGG       | 1140 |
| O | y | 1794 | tgagcaagtacacccgtgttgaaatgagaatgcactcctccasgtgctgtgcatalgacagtgtg    | 1853 |
| D | b | 1141 | TGGACAGTGAACACGGTGTGGAATGAGATGCATCTCCACATGCTGTGTCATGCGCAGTAGGG       | 1200 |
| O | y | 1854 | gctgcttcgtcgtgagactgtgctcttaaggtgtgctgtcagaagaagctcaagaatgatttcga    | 1913 |
| D | b | 1201 | GCTGCTCGCTGGACTGTGGCTTCAGAGTGGCTTCAGAGAGCTCAAGAAATGGATTGGCA          | 1260 |
| O | y | 1914 | tcaatcgcccccgagagacacacgcgcgaagaatcacaagcaatgagattctgctcttcca        | 1973 |
| D | b | 1261 | TCAATCCGGCCCCCAGAGACACACACGCGGAGAAATCCACAGCATGGAGATTCTGCTCTTCA       | 1320 |
| O | y | 1974 | actctgtagccatcaaccgcaaaaactcctcaagcagaagtgtgaacgttgggacaagctccta     | 2033 |
| D | b | 1321 | ACTCTGTAGCATATCAGCGCAAAATCTCCTACAGCAAGTTGAACGTGGGCGAAGTGTCTCA        | 1380 |
| O | y | 2034 | tctgtgagacttgggacatttcacacatgagcaatgagcaccagacagcgttctacaatgacctc    | 2093 |
| D | b | 1381 | TCTGTGAGCTGGGAACTTACACATTCGGCAATGTGCACCCAGAGGCGCTTATAATGTACCCCT      | 1440 |
| O | y | 2094 | ctgtgacctacaatctctctgcatcgctatgaacaagggaactcttctcaagctctctggg        | 2153 |
| D | b | 1441 | CTGTGCTCTATACATTCTCTCTCATTCGCTATGTACACAGGGAACTTTTTCAGAGCTTGGGG       | 1500 |
| O | y | 2154 | ctctctgaagaagttctgtgtgagagacacagcgttgggtgatacaatgtgaacgttgcatalgaaag | 2213 |
| D | b | 1501 | CTCTCTGAAGAGGTGTGGTGGAGAGACAGAGCGTGGGGTATCAATGTGAACGTGGCATGGACAG     | 1560 |
| O | y | 2214 | gaaggtgtggaaccccccaattggaagaagtgaagtacctcaagctctcaaggacagtgtgtga     | 2273 |
| D | b | 1561 | GAGGTGTGGACCCCCCATTTGGAGAGCTGTGAGTACTTACACCTTCACAGGACAGTGTGA         | 1620 |
| O | y | 2274 | tgcccatgtcccaagagttctcaacsfyabtggtcttaagctcccgccgggtttgaatgtgtg      | 2333 |
| D | b | 1621 | TGCCATTGCCCCACAGAGTTCTTCACTGATGTGTGCTCTTATGTCTCCCGCGGGTTTGTATGCTG    | 1680 |
| O | y | 2334 | ttgaagaagacatctgtctcctcttggtgtgctactctgtcaacgcgcagatgttltgtgcaact    | 2393 |
| D | b | 1681 | TTGAAGAGACATGTGTGTCCTCTGGGTGGCTACGTGTGTCACGCGCAGATGTTTTGGCCACT       | 1740 |
| O | y | 2394 | tgaccagagcagctgatatacctctgagcaagggtgccgggtgtgtctgaccttggaaggagcc     | 2453 |
| D | b | 1741 | TGACCAAGCAGCTGATATGACCTCTGGCAGGGGGCGCGGTGTGCTGTGGCCCTTGGCAGGAGGCC    | 1800 |
| O | y | 2454 | atgacttggaacgcacatctgtagatcgtctcgaagattgtgtctcggctctgtcactgattaa     | 2513 |
| D | b | 1801 | ATGACTTGAACCGCCCATCTGTGATGTGCTCTTGAGGCTTGTGTCTGTGCTCTGTGCTCACTGTAG   | 1860 |
| O | y | 2514 | agctgcagaccttgatataagcagctcttgacaaagaagccaaactcaacgacagtgtgcga       | 2573 |
| D | b | 1861 | AGCTGCAGCCCTTGTGATGAGGCAAGTCTTGTGCACAAAGCCCAACATCAACGCAGTGGCCA       | 1920 |
| O | y | 2574 | cgctagagaaagtcatcagatattccagagcaaaacacttgaagctgtgtgtgcagaagtctgcg    | 2633 |
| D | b | 1921 | CGCTAGAGAAAGTCAATGAGATCCACAGCAAAACACTGAGCTGTGTGCAGAAATGTGCCG         | 1980 |
| O | y | 2634 | ctgtgtctgggcccgtctccctctgcgggggccccaaagcaggtgataagccgaagaagccga-aatg | 2692 |
| D | b | 1981 | CTGTGTCGGGGCCGATCCCTCTGTCCGAGAGGCCCAAGCAAGTGTAGACCCAGAGAGCGCCGAGATTG | 2040 |

| QY | 2693 | tgaaagcgaatggccttgccttctgtgtgtgggagcgaataaggcccaagctgagggagccgg  | 2753 |
|----|------|--|------|
| Db | 2041 | tgaaacgcgcattggcccttgccttctgcgtgagggccagcagcgcacagccttgggcagccgg | 2100 |
| QY | 2753 | aacacagccccagcgccgagagagcccaatgagcagcagcctgcctgttaagcccg         | 2812 |
| Db | 2101 | aacacacccccagcccgccgagagagcccaatgagcagcagcctgcctgttaagcccg       | 2160 |
| QY | 2813 | gccccatcccttgggcttaccactgtgatlttgttattttctatataaaca              | 2872 |
| Db | 2161 | gcccccattccctctgggcttaccactgtgatlttgttattttctatataaaca           | 2220 |
| QY | 2873 | aagttaaat  | 2884 |
| Db | 2221 | aagtcacacatt   | 2232 |

|            |  |            |      |                 |
|------------|--|------------|------|-----------------|
| RESULT     | 6  |            |      |                 |
| AF132607   |  |            |      |                 |
| LOCUS      | AF132607   | 3255 bp    | mRNA | PRI 06-MAY-1999 |
| DEFINITION | Homo sapiens histone deacetylase 4 mRNA, complete cds.   |            |      |                 |
| ACCESSION  | AF132607   |            |      |                 |
| VERSION    | AF132607.1   | GI:4754906 |      |                 |
| KEYWORDS   |  |            |      |                 |
| SOURCE     | human.   |            |      |                 |
| ORGANISM   | Homo sapiens   |            |      |                 |
|            | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  |            |      |                 |
|            | Euthelia; Primates; Catarrhini; Homnidae; Homo.  |            |      |                 |
| REFERENCE  | 1 (bases 1 to 3255)  |            |      |                 |
| AUTHORS    | Grozinger,C.M., Hassig,C.A. and Schreiber,S.L.   |            |      |                 |
| TITLE      | Three proteins define a class of human histone deacetylases related to yeast Hda1p                                     |            |      |                 |
| JOURNAL    | Proc. Natl. Acad. Sci. U.S.A. 96 (9), 4868-4873 (1999)   |            |      |                 |
| MEDLINE    | 99238449   |            |      |                 |
| REFERENCE  | 2 (bases 1 to 3255)  |            |      |                 |
| AUTHORS    | Grozinger,C.M., Hassig,C.A. and Schreiber,S.L.   |            |      |                 |
| TITLE      | Direct Submission  |            |      |                 |
| JOURNAL    | Submitted (03-MAR-1999) Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA |            |      |                 |
| FEATURES   | Location/Qualifiers  |            |      |                 |

```

CDS
      1..3255
      /note="HDAC4; similar to Saccharomyces cerevisiae Hdacp"
      /codon_start=1
      /product="histone deacetylase 4"
      /protein_id="MAD29046.1"
      /db_xref="GI:4754907"
      /translation="MSSQSHPDGLSGDQPEYELLNPRVNHMSTVYATPLQVAPR
      SAMPMDLDHGFSLPAEERPAERQQLDQDELLAKQKQQLQRLQLLEFQKHQDLSSR
      QHEQALHEHNRKQOQEMLAMKHQOELLHORLEHNRKQOELEKHQKQDLQNLKKE
      GKSAQVASTVEVMKKLQEFVFNKKALAHNRNLNHCISDPRKMYKQTHSSLDQSSPPQ
      SVYSTVNHVPIVGYAKDKDPDLKPTKSTSEPMPLKRSKRLQKVARRRSSPLRRDGPV
      VTLLKRRPLDYVTSACSSAPSGSGSPSSNNSSGVSANGLNPAVPSYIPAEISLHRLV
      ARGSAAPRLPYTSPSLPNTLLGLPAIGSPASGAGOODTELLTLPALQRLSLRPGNH
      LTYLTYLSTPLERDGGAHSPRLQHMVLELDPPAAPRLVTGGLALQASLVGADRRV
      PSILHKLRHRRPLRGTHQASPLPQNNQALQHLVLYQHOHPLFEKHQKQOEQDQNLKIL
      PKSEPARQPESHNEPEETEELREHQAALLDERYLDRLRGQKKAHQAAQVQYKQETESD
      EEEAREPREVEPQQRQSEQLLETRQOALLLEQQRHQLNRYQASMSAAGIPVAFGH
      RPLRSQSSPASATFFPVYQDEPRTPKPFETTGILVYDRLMLHOCQSGSSSHPEAGRI
      QSILNROETGLRGKQCICIRGRATLEELQVSEATLLLYGTNPLNRQKLDISKLL
      SLASVRLPLCGGVGQVDSPTIMNVEVSAGAARLAVERLVEFVATGLCKNGAVVVR
      PCHHHAESTPMGFCYTNISVAVALAKLLQQLRSLSKILLVMDVHHNGCTQOAFSDPS
      VLVMSLHRDYDGNFPGSGADDEVGTGPGVFNNMAKFTGGLDPPDAETLAFRRV
      VMVIASEFAPDVLVVSFGFADVEGHPRLGCVNFAFEYLTQLMGLAGGRATVLL
      EGGHDIATICDASACVSLALGNLELDLPKEVLOQRNANANVRMEKVEMLHSHYMC
      LQTTSTAGAGSLLEAQTCEBDEAEETVTAASLSTGVCAERKDEPEEMEEPRPL"
BASE COUNT
      701 a      1070 c      988 g      496 t
ORIGIN

```





Db 2599 GATGGAACTTCTCCAGGAGCGGGCTCTGATGAGTGGGACAGAGCGCGCTG 2658  
Qy 2186 ggagcaaatggaacgctggagatgagagagtgatgagagagagagagagagag 2245  
Db 2659 GGTTCACAGCAACAGATGCTTACCGGCGGCTGGAGCCCGCCACAGGAGCGTGA 2718  
Qy 2246 tacctaacagccttcagagacagtgatgagccattgcccagagatctcaccatg 2305  
Db 2719 TACTTGGCGGCTTCAACAGGTTGATGCGGATGCGACGAGTTTGGCCGAGATG 2778  
Qy 2306 gtccatgctcccgagagtgatgagagagagagagagagagagagagagag 2365  
Db 2779 GTGTGCTGTCATCAGGCTTCGATGCGGAGGAGGAGGAGGAGGAGGAGGAG 2838  
Qy 2366 tctgtaacccagagatgtttgagccattgacagagagagagagagagagag 2425  
Db 2839 AACCTTCCGCAATCTTGGGTACCTGACGAGGAGGAGGAGGAGGAGGAGGAG 2898  
Qy 2426 cggatgagtgagcctgagagagagagagagagagagagagagagagagag 2485  
Db 2899 CGGATTTCTCTGCGCTTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2958  
Qy 2486 gcttgatgctcagagagagagagagagagagagagagagagagagagag 2545  
Db 2959 GATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3018  
Qy 2546 caaagagcccaatcaacagagagagagagagagagagagagagagagagag 2605  
Db 3019 CAAGAGCCCAATGCAACAGGCTGCTTCCATGAGAGAGAGAGAGAGAGAGAG 3078  
Qy 2606 cactgag 2665  
Db 3079 TACTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3138  
Qy 2666 gcaagtgag 2724  
Db 3139 ACTTGCAG 3198  
Qy 2725 gaacagagcccaagagagagagagagagagagagagagagagagagag 2788  
Db 3199 AAGCCCGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3232

RESULT 7  
LOCUS AB006626 8459 bp mRNA PRI 25-DEC-1999  
DEFINITION Homo sapiens mRNA for KIAA0288 protein, partial cds.  
ACCESSION AB006626  
VERSION AB006626.2 GI:6635126  
KEYWORDS KIAA0288 protein; KIAA0288.  
SOURCE Homo sapiens Brain cDNA to mRNA, clone:lib:psport 1 clone:ha061161.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 8459)  
Ohtsuka, O., Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N.  
and Nomura, N.  
Prediction of the coding sequences of unidentified human genes  
2 (bases 1 to 8459)  
Ohtsuka, O., Nagase, T., Ishikawa, K., Nakajima, D., Ohira, M., Seki, N.  
and Nomura, N.  
Direct Submission  
JOURNAL Submitted (20-NOV-1997) to the DDBJ/EMBL/GenBank databases, Nobuo  
Nomura, Kazusa DNA Research Institute, Gene Structure; 1532-3  
Yana, K. Saito, Chiba 292, Japan (E-mail:cdna@kazusa.or.jp,  
URL: http://www.kazusa.or.jp, Tel: +81-438-52-3930,  
Fax: +81-438-52-3931)  
COMMENT On Dec 25, 1999 this sequence version replaced gi:2564323.  
Sequence updated (20-Dec-1999).  
FEATURES  
source 1..8459

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="ha061161"  
/clone\_lib="psport 1"  
/tissue\_type="Brain"  
754..4047  
/gene="KIAA0288"  
<754..4047  
/gene="KIAA0288"  
/note="Start codon is not identified, ha06116 cDNA clone  
for KIAA0288 has a nucleotide substitution at the position  
1135 ( C -> T ) of the sequence of KIAA0288"  
/codon\_start=1  
/product="KIAA0288 protein"  
/protein\_id="BAA22957.2"  
/db\_xref="GI:6635127"  
/translation="AISNHLKENTLLAMSSQHPDLSGRDPEVLLNARVNHMST  
VDVATLPLQVAPSAVPMRLDHPFLPAEPALRQDLQDLALRQKQIQRQIL  
IAEFORHEDLSRQHEAOIHETIKQOEMLAMHQDELHQRKLEHREDELEKOH  
REQLQDLKREKESAVASTEVKKMLQKALALANRNLNHCISDPRVYWG  
TQSSLDSSPPQSGVSTYNNHPLGMVAKDPPRLKRTASRNLTLSRLKQVNR  
RSSPLLRKRDGPVYVYALKRRLDVTLSACSAGSSPSSPNNSSGVSANGLAPV  
STPASTSLARLVARGSAAPLPLVYSPSLNLTLLPATGSPAGTAGQDTERLPL  
ALQRLSLPFGTHLPLVLTSPLEKRGGAHSLQHMVLEQPPAPAPLVGLALP  
LHQSLSVADRVSPSIHLKROHRLDGRTOASAPLQNAOLHVLVIOOHOLEKHO  
OFQOOLQMKRIIPKSEPAROPESHPEETEELREHQAOLHDLPLRLPGKEHAHO  
AGVQKOEPIESDEEAEPRVPEPQROPSDELEPFOALILEORIHOLNVCAS  
MPAAGTIPVSGGHRPLSRAOSSPASAPVPSVSGPPTGGLVYDPTLMKHOCT  
GSSSPHPRHARLIQSTVSRILQETLKGKCEITGRKATLELDTVSEATHTLYKIP  
LNQRKLSRHLKLSLASVFLRPLCGGVGVDSDTINMVSAGARLAVGCVLVEKYV  
ATGELNENFAVVRPGRHAAEESTPMGFCYFNVAVALKLOQLSVSKTLLVDMVNH  
GNGTOAFYSDPVLVYMSLHRYDDGNFPGSGAPDVGPGVFNMAFTGGLDP  
MGDAEVLAAFRVYVPLASPEAPDVVLSGFPVAGHPPLGGLYMLSRACFGYLRKQ  
LMGLAGRIYVLAEGHDLATQDASEACYSALLGNELDLPKRVLOQPNNAVSM  
EKYMEIHSKYWRCLQRTTSTAGRSLLIEAQTCEMEBEAYVTAAASLSVGVKPRKRPDE  
EPMEBEPLP"

BASE COUNT 1852 a 2375 c 2338 g 1894 t  
ORIGIN

Query Match 33.1%; Score 953.6; DB 9; Length 8459;  
Best Local Similarity 62.9%; Pred. No. 5,4e-187;  
Matches 1720; Conservative 0; Mismatches 919; Indels 95; Gaps 12;

Qy 43 cctcaaccattccctccacagagagagagagagagagagagagagagagagag 102  
Db 1368 CCACTGATTTTCCAGCCACCTCTGCTGATGAGGAGAGAGAGAGAGAGAGAG 1427  
Qy 103 ccaagagtcctccctccacagagagagagagagagagagagagagagagag 162  
Db 1428 CCAGAGTTTCTCCAGCCAGAGC-----GGAGTCTCGACCTCTATTAACCGGGT 1478  
Qy 163 gacctgag 222  
Db 1479 CCTGGAGATGTACACACCCAAAGATGACTTCCCTTGAAGAAACACCTCTGAACGAA 1538  
Qy 223 ctgtaagagtgagtgagagagagagagagagagagagagagagagagagag 282  
Db 1539 TCTGAAATTTACGCTCCAGGCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1598  
Qy 283 gcgtcgaag 342  
Db 1599 ACGCAGGAG 1658  
Qy 343 tgcggag 402  
Db 1658 -----ACTCCGGGTGTCACAGGCGCCACAGGCTCCGAGCCAGCTCAC 1700  
Qy 403 -----caacagctccacagagagagagagagagagagagagagagagagag 459  
Db 1701 CAACACAGCTCCGCGGAGCGTCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1760  
Qy 460 cccacag 519



|    |      |  |      |
|----|------|--|------|
| Db | 1761 | CCCCGGAGAGACGAGTTTGGCGCACAGACTTGGGACACGAGAAGGCTTGCGCGCTCACT          | 1820 |
| QY | 520  | cagcctctacacgctctccctctctgtccccaacatctccctcagggtcgacgagcaggtac       | 579  |
| Db | 1821 | TCCCTTACGACATCGGCATCTCTTGCCCAACATACAGCTGGGCGTCCCTGCCACGGCC-          | 1880 |
| QY | 580  | tgtaaccaactacaacctcaactcagctccctcccgaaagctgtctgcacagacgaaagggccgagag | 639  |
| Db | 1880 | -----CCTGTGGGGGACAGCGGGGCGGACAGAGACACCGAGAG                          | 1916 |
| QY | 640  | gcagagccctccagctccctctgagcagaggtgagacgctgacccggcaagttcatgagacatc     | 699  |
| Db | 1917 | ACTGACCCCTTCCCGCCCTCCACGAGAGAGGCTCTCCCTTTTCCCGGCACCCACTTACCTCC       | 1976 |
| QY | 700  | ctctatctccgtcgtcgtctgtctggtgagtgacgtggagggcgagcgagagccccaagcgac      | 759  |
| Db | 1977 | CTA-----CCTGAGACATCTGCCCTTTGGAGCGGGAGCGAGGGGCA--GCGCA                | 2021 |
| QY | 760  | tgacctccctcgtcagcagatgtctgtgtgtcgtgagcagcgccgagcagcagagaccctcat      | 819  |
| Db | 2022 | CAGCCCTTCTTGACGACATATGGCTTACTGGAGCAGCACCGGACCAAGCACCCCTCGT           | 2081 |
| QY | 820  | tgtctgtccactccaagcgagctccccaatgtagcgggtgtagcgtgtggtccacagcat         | 879  |
| Db | 2082 | CACAGGCGCTGGGAGACATGTGCGCCCTCCACAGACAGTCTGTGTGGTGGAGACCGGGTGTG       | 2141 |
| QY | 880  | gcgagcgtgtagcagagctcccgcgagatcagcgccctgagccgcatctagctccacgct         | 939  |
| Db | 2142 | CCCTCATTCACAGATGGTGTGGGACGACCGCCACTGGGGGCGACCCATGTGGCCCGCT           | 2201 |
| QY | 940  | gcgcagagatccccaagcgccctcagcagcagctgtgtcatgtacacaaacagcagcagctct      | 999  |
| Db | 2202 | GCCCCAAGAGGCGCAGGCTGTGGACGACCTGTGCATCCAGCAGCATCAGCATCTTCT            | 2261 |
| QY | 1000 | ggagaaagcaga-----agcagcagcagctacagctgtggcagaatctccaa                 | 1047 |
| Db | 2262 | GGAGAAACACAGACAGCAGTTCACGACGACGACGACATGTGAACAAAGATCATCTCCCA          | 2321 |
| QY | 1048 | gacagggagatgcccagggagcgcacacacccccttgagagagagagggagagctgtac          | 1107 |
| Db | 2322 | GCCAAAGCAGAGCAGCCCGGACGCGGAGAACCCCGGAGGAGAGGAGAGAACTCG               | 2381 |
| QY | 1108 | ggagcagcagagaggtctctgtctggggagagggagcctgcagcaagcagccggagaggtccac     | 1167 |
| Db | 2382 | TGACACACC---AGGCTGTGGTGGACGACGACCCCTACCTGGAACCGGCTGCGGGCACAAGA       | 2438 |
| QY | 1168 | agagaggtgagagacacagaaagacctgtgagagagagagcgaagaaagagatgtggggagga      | 1227 |
| Db | 2439 | GGGCCACGACACAGGCGCGCTGTGACGTGAAGCAGGAGGCCATTGGAGAGGATGAGAGAGA        | 2498 |
| QY | 1228 | ggagagagatgtcatccaggttaagagcagagagggcgagatgtgtctgtcggagaggggc        | 1287 |
| Db | 2499 | GGCAGACCCCCACGSGGAGGTGAAGCGGGGCCAGCCGACCCCACTGAGCAGAACTCT            | 2558 |
| QY | 1288 | cgaattggaaga--gcctggtgtctgatacgaaaactgttctcagatgagcccaacgctg         | 1345 |
| Db | 2559 | CTTGACAGACAGACCTCTCTGCTGGAGCAGCAGCGGATCCACACAGCTGAGGAAC-----         | 2614 |
| QY | 1346 | caaccttgcaggtgttaacaagcgcgcctcagcctgtgagcactgtgtcccaaccaagcctg       | 1405 |
| Db | 2614 | TACCAAGCGTCATATGAGGAGCGCGCGCATGCCGTCCTTGGCGGGCACAGGCTCTG             | 2673 |
| QY | 1406 | ggcgcgtaccaatctctccctgtctgcacctgtggggagatgaagaaaccccccagacaacc       | 1465 |
| Db | 2674 | TCCCGGGGCGAGTCTTACCCCGGCTGTGCACCTTCCCGGTCTGTGTGAGGAGGCCCC            | 2733 |
| QY | 1466 | gtcaagacacctctctacacaaagtgtgtctacagacgttcatgtctaaagacacagctgc        | 1525 |
| Db | 2734 | ACCAAGCGGAGGTTCAAGAGAGGCGCTGTATGACACCGCTATCTGAAGCACAGTGCG            | 2793 |
| QY | 1526 | atgtcggggaacacacagctgcacccctgagcatgtctgcgcggtaccagatcctgtctc         | 1585 |

|    |      |  |      |
|----|------|--|------|
| Dd | 2794 | ACCTGCGGGAGTACACGACACGCCACCCCGAGCACCCCGGAGAGATCCAGACATCTGCTCC      | 2853 |
| Qy | 1586 | cgagctcagagagacaagcgctcgcttagcaagtgcgagcgagtcacgagtlcgcgaaacacag   | 1645 |
| Dd | 2854 | CGCCTGAGAGAGAGAGGGGCTCTCGGGGGCAATGGAGATGATCTCGGGAGACGCAAGGCCAC     | 2913 |
| Qy | 1646 | ctagaatgagatcacaagacagtgcacactgtaatacacacacactgctctatgtagacagctcc  | 1705 |
| Dd | 2914 | CTGGAGAGGCTACAGACGGGTGCACATCGGAAGACCCACACCTCTGTATATGGACAGAACCC     | 2973 |
| Qy | 1706 | ctcaacgcgagaaagcgaagcagaagaagtgtgctgctccatcaagccagaagaagtgtat      | 1765 |
| Dd | 2974 | CTCAACCGGGCAAAACGTGGACAGTAAGAAACTTATAGCTCTCCTCGCC--TCGCTGTC        | 3030 |
| Qy | 1766 | gcctgtgctgctcttgagggacatcgaggtgtgacagtgcacgcgtgtgatatgatatgac      | 1825 |
| Dd | 3031 | GTCCGGGCTCCCTTCCGGTGGGTGTGGGGTGGACAGTGCACACCATATGAAACAGAGTGCAC     | 3090 |
| Qy | 1826 | tctctcagctgctgtgcagatgtagcagttgagctgtgcctgcgtagagctggtcctcaagtgct  | 1885 |
| Dd | 3091 | TCGGCGGGGGGACGCCCTCGGCTGTGGCTGCTGCTGTGAGACTGGTCTTCAAGTGTGCC        | 3150 |
| Qy | 1886 | gcagagagagctcgaagaatgtgattgtccatcatccgacccccagagacaacgcgcgagaa     | 1945 |
| Dd | 3151 | ACAGGGGAGCTGGAAGATGTGCTTGTGTGTGTCGCCGCCCTCGAGACCATCTCGGAGAG        | 3210 |
| Qy | 1946 | tccacagccatggagattctgtcctctcaactctgtgatactgataccgacaanaactccaaag   | 2005 |
| Dd | 3211 | AGCACGCCCATGGGCTTTTGCTACTTCACCTCGGTGGCCGTGGACCAAGCTTTGTGAG         | 3270 |
| Qy | 2006 | cagaagctgaaacgtgaggacaaggtctccatctgttgaacttggagacattcaacatgtgcaatg | 2065 |
| Dd | 3271 | CAGAGGTTGACCGTAGACAGATCTCATGTGTGAGATGGAGAGTGCACCATGGAAGAACGG       | 3330 |
| Qy | 2066 | accacagcagcgctctcaaatgtgacccctgtgctctacatctctgcatactgatactgtac     | 2125 |
| Dd | 3331 | ACCCAGCAGGGCTTCTACACGACGACCTTAGCGTCTGTACATGTCCCTCCACGCGTACAC       | 3390 |
| Qy | 2126 | aacgggaactctcttcgaagctctgcgggctccctgaagaggtttgtgtgagagaccagcggt    | 2185 |
| Dd | 3391 | GATGGGAACCTTCTTCCAGGACGCGGGGCTCTGTATGAGGTGGGCACAGGGCCCGGCTGG       | 3450 |
| Qy | 2186 | gggtacaaatgtgaacgctgtgatatgtgacaagaggtgtgtgacccccccatttgaagacgttag | 2245 |
| Dd | 3451 | GTTTCAACGTCACATGAGGCTTCAACCGGGGGCTGTGACCCCCCATGGGAAACGCTGAG        | 3510 |
| Qy | 2246 | taccttacacgctcagagacagtgtgtatgtgccattgtgccacgaagtctcacgtatgt       | 2305 |
| Dd | 3511 | TACTTGGGGGCTTCAACAAGCGTGTCTATGCCATGGCCAGGAGTTGCCCCGGATGTG          | 3570 |
| Qy | 2306 | gtctcagctctccgcgcgagttgtatgtcgtgtgaaggacatctgtctcctctgtgtgtgtac    | 2365 |
| Dd | 3571 | GTGCTGGGTATATGAGCTTTCAGTGCCTGTGAGGGGACCCACCCACCTCTTGTGGGGCTAC      | 3630 |
| Qy | 2366 | tctgtacacgcgcagaatgtttgtgcacactgtacacagcagctgatatgacccctgtgcaggggc | 2425 |
| Dd | 3631 | AACCTCTCCGCAATGCTCTTCGGGTACGTACACAACACACTGTATGGGCTCGGCTGGCGGC      | 3690 |
| Qy | 2426 | cgagtggtgtgctgacctgtgagagagggccatgacgtgacgcgcacatctgtatgtgctctgaa  | 2485 |
| Dd | 3691 | CGGATTGCTCGGGCTCTCGAGGGAGGGCACAGACTTACCGCATTTGGCAGCCCTCTCGGAA      | 3750 |
| Qy | 2486 | gctgtgtctgtgctgtgcctcagtgtaaaagctgtgcagcccttgatgtgaagcagctgtgag    | 2545 |
| Dd | 3751 | GCATGTGTCTGTGCTTGTGCTGGGAAACGACTTGTATCTCTCCACAAAAGGTTTACAG         | 3810 |
| Qy | 2546 | caaaagccacaacatcaacgcagctgtgccaacgctagagaaatgatactgagatctcagagcaaa | 2605 |
| Dd | 3811 | CAAAAGACCAATGCAAAACGCTGTCTCGTTCCATGTGANAANAATCATGAGATCCACAGCAAG    | 3870 |
| Qy | 2606 | cactgtgagctgtgtgcaagaagtctgcgcgctgtgtcgtgcccgtctccctgcgagggcccaa   | 2665 |
| Dd | 3871 | TACTGTGGCTCTCCGACGCCACAACTTCCACAGCGGGGCGTTCCTGTATGTGACGACTAG       | 3930 |

[illegible][illegible]

OY 2435 ctggcccttgagaggagccatctgaccccatctgtatgctctgtaaagctgtc 2494  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2288 CTGGCCTTGAGGGGTGGCCTAGACCTCACAGCCATCTGTGAGCGCTCTGAGCGCTGTGTG 2347  
OY 2495 tcgagctctgacagtgtaaaagctgcaagcccttgatgagagagctctgcaagaaagccc 2554  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2348 GCTGCTCTTCTGGGTAACAGAGGTGGATCCCTTTCAGAGAAGAGGTGGAACAGAAACCC 2407  
OY 2555 aacatcaacgagctgagccagctagagaagatcagagatccagagcaaacactgagagc 2614  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2408 AACCTCAATGCATCCGCTCTCTGAGAGCGCTGATCCGGGTGCACAGTAATACTAGGGGC 2467  
OY 2615 tgtgtgagaagagctgcgcgctgctgagccgctgctgagagggcccaagagagtgag 2674  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2468 TGCAATGCGCGCCTTGCGCTCTGTCCAGACTCTGAGGCTCTAGAGTCCAGGGGCTGAC 2527  
OY 2675 accgaagaagcgaatctgaacgccatgagcttgc 2710  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2528 AAAGAAGAGTGAAGGAGCACTAACCGCACTGGCGTCC 2563  
RESULT 9  
AF207749 2871 bp mRNA ROD 07-FEB-2000  
LOCUS AF207749 histone deacetylase 7 (HDAC7) mRNA, complete cds.  
DEFINITION AF207749  
ACCESSION AF207749.1 GI:6911183  
VERSION AF207749.1  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
REFERENCE 1 (bases 1 to 2871), Ordentlich, P. and Evans, R.M.  
AUTHORS Kao, H.Y., Downes, M.,  
TITLE Isolation of a novel histone deacetylase reveals that class I and  
class II deacetylases promote SMRT-mediated repression  
JOURNAL Genes Dev. 14 (1), 55-66 (2000)  
MEDLINE 20107033  
REFERENCE 2 (bases 1 to 2871)  
AUTHORS Kao, H.Y., Downes, M., Ordentlich, P. and Evans, R.M.  
TITLE Direct Submission  
JOURNAL Submitted (22-NOV-1999) Gene Expression Laboratory, The Salk  
Institute for Biological Studies, 10010 N. Torrey Pines Rd., La  
Jolla, CA 92037, USA  
FEATURES  
source  
1..2871  
Location/Qualifiers  
/organism="Mus musculus"  
/strain="B6:C57BL/6"  
/db\_xref="taxon:10090"  
/dev\_stage="day 14.5 embryo"  
1..2871  
/gene="HDAC7"  
1..2817  
/gene="HDAC7"  
/function="repress transcription"  
/note="class II histone deacetylase"  
/product="histone deacetylase 7"  
/codon\_start=1  
/protein\_id="AA01419.1"  
/db\_xref="GI:6911184"  
/translation="MHSFGACRALQPDPTGSGROPMDLVGQRPVTEPPERALLTL  
QHPQRHLRLAGLHDOORSAEPMRLSDPMPPELQGGQDELRQLLNKDKSKRA  
VASSVYKQKLAELYILKQOALERTVSPSPISPYRLIEDLDEGAARVLSFLP  
PSLPTPEPRPLKRYVSEPNLKLRYPKKSLERKNPLRKEBAPSLRRPAETLG  
DSSPSSSTPAGSCSSPNDEHGNPALGSDADDRTHSTGLGRGVLPNPAHFL  
HGLLEPAGGTLRLOPILLDPSVSHALMTYPGIGPLPHFAOPLTTERLSGS  
LHRLPRTREBPRLPSATSLAPLAPLOPRDRKLPHYOLIKPAISPORAKREBKR  
LRQIPSAEDLETJBGGVGMANDGLEHRESGRGPEGRGSIQLQHOQVPWEQHLA  
GRLSQSGPQSVLPIPLAQVGRPLSRQSSPAARVSMLSPEPTQOTVLSSETPAG  
LVYDSVMILKQSCGDNKSRHEHAGRIQISIRLQERGLRSCCLGRASALELDS  
VHSERHVLGTNLSRLKLDNGKLTGLLQRTFVMLPCGVGVDTPTINNELHSSNA  
ARMAAGSVMDLAKFVAARELKNKNGFAVVRPPGHADHSTAMGFCFNSVALACRLQOH

BASE COUNT 585 a 946 c 801 g 539 t  
ORIGIN  
Query Match 21.7%, Score 625.8; DB 12; Length 2871;  
Best Local Similarity 67.6%; Pred. No. 2,56-119;  
Matches 908; Conservative 0; Mismatches 432; Indels 3; Gaps 2;  
GKASKLILVMDVHHNGTOOTFEYODPSVLYISLHRHDDGNFPFGSAGVDEVTASGE  
GFNNVAMAGLDPMDPEYLAERFRIVVMDIAREFAVDIYVAGDAEAGHPAPIG  
GYHSACFGYMTQOLMNLNLAGAVYLLBEGHDIETATCDSEAKVALLKRXDPLSE  
ESMRQKPNLSAIRLEAVRYHRVRYWCMQMLASCPDPSWLPRLVPGADAEEAVYALAS  
LSVGLIADRSERLVEEPEPMN."  
OY 1379 ctggcacaatggtgccccaccagaagcccttggtggccgtaccacaatctccctgcgtg -ccccgg 1437  
Db 1413 CCGGCGCTGTCCAGAACCCAGACTCTTCCCGACAGCACACTGTTTCCATGTTGAGACAGA 1472  
OY 1438 gggcattgaagaacccccagagaccacccgtcaagcactcttcacacaaagtgtgtcta 1497  
Db 1473 GCCACACTGTGACAGACCCCAAGTCTTCAACAGCTCAGAGACACACTGCTACAGGGCTGTCTA 1532  
OY 1498 cgaacgltcatgctaaagcaccagtgcatgtgcgggaacacacacgctgcacccctgagca 1557  
Db 1533 TGACTCGGTGATGCTGAACAACCAATGTTCTGTGAGAGAACAAACAGACATCCCGAGCA 1592  
OY 1558 tgcgtgcccagatccagagcatctgtgtcccgctgagagaagagcctgcttagaagt 1617  
Db 1593 TGCAGGCGCATCGAGACATCTGTGTCCGGCTGAGAGACGGGGTCTCCGACGCACTG 1652  
OY 1618 cgaagcgatccgaagctgcgaagcagcctagatagatccaagaagctgcactcgaata 1677  
Db 1653 TGAGTGTCTCCGAGGCGCCAAAGAGCTTCCCTAGAGAGAGCTGCAGTACGCTCTGAACG 1712  
OY 1678 ccacacccctgctctatgagacagtcctccctcaacccgycagaagctagacagcaagaat 1737  
Db 1713 GCACGTGCTCTTTCACGACAGAACCCACTAGCCGCTCAAACTGATGAACGGGAACT 1772  
OY 1738 gctcgtgtcccatccagcagaagatgtatgtcgtgtgctcctgtgtgagagcaccgagtg 1797  
Db 1773 TACAGGACTCTCGGACGACGCGAGCTTTGTGATGCTCTGTGGCGGGGTTGGGTGGA 1832  
OY 1798 cagtgaacagctgtggaatgagatgactctcctccagctgtctgagcagtgagtgagctg 1857  
Db 1833 TACTGACACCATCTGGAACGAGACTCATTTCTCCATGACAGCCGCTTGCGTGGGAG 1892  
OY 1858 cctgctgagcttggtccttcaagtgctgctgcagagagctccaagattggtccatcat 1917  
Db 1893 TGTACACGACCTTGCTTCAAAAGTAGCTTCCCGAAGAGCTGAAGAATGGCTTGTGCTG 1952  
OY 1918 ccggaccccaaggacaacacgcgcgaggaatccaagacatggagatctgtctcttaactc 1977  
Db 1953 GCGAGCCCCGGGACACCATGAGATGAGATCAACCCATGAGGCTTCTCTTCAACCTC 2012  
OY 1978 tgtagcacaacgcgaaaactcctcacaagcagaagtgtgaagctgagcagtgatccatcgt 2037  
Db 2013 CGTGGCATTCGCTGCGCCACAGACTTACAGCAACCGGCAAGCCACAGATCTCATTTGT 2072  
OY 2038 ggaactggagacattcaacattgacattgacccagcagtgcttctctaaattgaacccctctgt 2097  
Db 2073 TGACTGGGATGTTCAACATGGGAAGGACAGACAGAGACTTTCTACAGAACCCAGTGT 2132  
OY 2098 gctctaaatctcttgcattgcattatgacaagaggaaactcttccagagctctgaggtcc 2157  
Db 2133 GCTTACATTTTCCCTTCAATCGCAAGAGAGAGCAACTTCTTCCAGAGAGGCGCGGT 2192  
OY 2158 tgaagaaggttggttgaaggaacagagcggtgggtatacaattgaacctgagatggagagaag 2217  
Db 2193 GGATGAGGTGGGAAGTCCACAGTGGCGAGAGGCTTCAATGTCAACGATGGCTTGAGGG 2252  
OY 2218 tgtgagcccccatctgagacgltggaatlaacttacaagccttcaagaagatgtgtatgct 2277  
Db 2253 CTTGATATCACCAACAGGGGAGATCCTGAGTACCTGCTGCTTCAAGAGATAGTGATATGCC 2312

|            |  |   |                     |
|------------|--|---|---------------------|
| Oy         | 2278   | cattggccacgaatcttcaaccgtagtggtcctaagtcctccgcggcggtttgatgaccttta   | 2337                |
| Db         | 2313   | CATTGCCCGAGAGTGTGCTCCAGACCTGGCTTCGTGTGTGTGGGTTTGATGACTGCGA        | 2372                |
| Oy         | 2338   | aggacatctgtctccctcgtggtgtactctbtcaaccgccaaatgttttgcaccattgac      | 2397                |
| Db         | 2373   | GGGTCAACCACGCCCCGGCTGGGTGGGTACCATTTTCTGTGCAAAATGTTTTGGGTACATGAC   | 2432                |
| Oy         | 2398   | caggcagcttgatgaccttgcagaaggccggagtgtctgtgcctcgtgaaggacata         | 2457                |
| Db         | 2433   | GCAGCAGTTGATGAACTTGGCAGGAGGGGCCCTGGTGTTGGCTTAGAGAGGTGGACATCA      | 2492                |
| Oy         | 2458   | cttaaccgccaatcttgtatgtcctctgaagactgtgtctcgcgtctgcatagttaagct      | 2517                |
| Db         | 2493   | CCTCACGGCCCATCTGTGTATGCTTCGGAGGCTGTGTAGCTGCTCTTCTTG6GACAAGAAT     | 2552                |
| Oy         | 2518   | gcagcccttgatgtgaggaagctttbcagcaaaaagcccacaatcatcaacgacagtcacgct   | 2577                |
| Db         | 2553   | GGACCCCCTTTTCAGAAGAAAGCTGAAAACAGAAACCAAACCTCAGTGGCATCCGGTCGCT     | 2612                |
| Oy         | 2578   | agagaaagtcattcgagatccagagcaaacacttgcagcttgtgtgcagaagtctgcgcgtg    | 2637                |
| Db         | 2613   | GGAGCTGTGTGTACAGGGTGCCACAGGAAATACTGGGGGCTGTCAATGCAGCGCTTGCCCTCG   | 2672                |
| Oy         | 2638   | tctggscgsgtccccctgcgagggcccaaagcaagtgaagccgaagaagccgaatatgaac     | 2697                |
| Db         | 2673   | TTCACACTCTCTGTGGTATACCACAGATGCCGGAGACTGTATGCAGAAATGG - AACCCGTACC | 2730                |
| Oy         | 2698   | gcacatgacctgtctgtgtgtgtg 2720                                     |                     |
| Db         | 2731   | GCGGTGGCATCCCTTTCTGTGGG 2753                                      |                     |
| RESULT 10  |  |   |                     |
| LOCUS      | AC004150   | 120766 bp   | DNA PRI 17-APR-1999 |
| DEFINITION | Homo sapiens chromosome 17, clone hRRC.97L_F_3, complete sequence.   |   |                     |
| ACCESSION  | AC004150   |   |                     |
| VERSION    | AC004150.8   | GI:4585948  |                     |
| KEYWORDS   | HTG.   |   |                     |
| SOURCE     | human.   |   |                     |
| ORGANISM   | Homo sapiens   |   |                     |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;        |   |                     |
| AUTHORS    | Eutheria; Primates; Catarrhini; Hominiidae; Homo.                    |   |                     |
| TITLE      | 1 (bases 1 to 120766)  |   |                     |
| JOURNAL    | Birren,B., Linton,L., Nusbaum,C. and Lander,E.                       |   |                     |
| REFERENCE  | Homo sapiens chromosome 17, clone hRRC.97L_F_3                       |   |                     |
| AUTHORS    | Unpublished  |   |                     |
|            | 2 (bases 1 to 120766)  |   |                     |
|            | Birren,B., Pasman,K., McKernan,K., Nusbaum,C., Richardson,P.,        |   |                     |
|            | Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckery,R.,     |   |                     |
|            | Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P.,      |   |                     |
|            | Daily,M.J., Depayre,E., Devon,K., Dewar,K., Donelan,L., Durette,B.,  |   |                     |
|            | Flemadi,S., Ferreira,P., Forrest,C., Funke,R., Gage,D., Gardyna,S.,  |   |                     |
|            | Genschmer,S., Gerlgerly,K., Gilmarlin,T., Gray,D., Hagos,B.,         |   |                     |
|            | Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Linton,L.,   |   |                     |
|            | McKenzie,J., Marguis,N., McEwan,P., McCurk,A., Meldrum,J.,           |   |                     |
|            | Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T.,  |   |                     |
|            | Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D.,          |   |                     |
|            | Rollins,G., Rossello,R., Roy,A., Shyam,R., Soohoo,S.,                |   |                     |
|            | Strange-Thomann,N., Stilwell,J., Stone,C., Strickland,C., Sydney,K., |   |                     |
|            | Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y.,          |   |                     |
|            | Ye,W.J., Zemseva,I., Zhao,J. and Zody,M.                             |   |                     |
| TITLE      | Direct Submission  |   |                     |
| JOURNAL    | Submitted (16-FEB-1998) Whitehead Institute/MIT Center for Genome    |   |                     |
| REFERENCE  | Research, 320 Charles Street, Cambridge, MA 02141, USA               |   |                     |
| AUTHORS    | 3 (bases 1 to 120766)  |   |                     |
|            | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  |   |                     |
|            | Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,       |   |                     |
|            | Castle,A., Cerny,J., Collangelo,M., Collins,C., Collymore,A.,        |   |                     |
|            | Cooke,P., DeRellano,K., Depayre,E., Devon,K., Dewar,K.,              |   |                     |
|            | Donelan,L., Doyle,M., Ferreira,P., Fitzhugh,W., Forrest,C.,          |   |                     |
|            | Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,     |   |                     |

|          |  |
|----------|--|
| TITLE    | Hagos, B., Heatfoot, A., Horton, L., Howland, J.C., Jones, C., Kann, L.,   |
| JOURNAL  | Karatas, A., Lehoczek, J., Liu, C., Locke, K., MacDonald, P.,  |
|          | Martius, N., McSwan, P., McGuck, A., McKernan, K., McLaughlin, J.,   |
|          | Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,   |
|          | Naylor, J., Ntloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,   |
|          | Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,  |
|          | Strange-Rhmann, N., Stojanovic, N., Stone, C., Subramanian, A.,  |
|          | Wesley, S., Torruella Miller, I., Vassiliev, H., Vo, A., Wagner, A.,   |
|          | Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, W.   |
| COMMENT  | Direct Submission<br>Submitted (17-APR-1999) Whitehead Institute/MIT Center for Genome<br>Research, 320 Charles Street, Cambridge, MA 02141, USA<br>On Apr 16, 1999 this sequence version replaced gi:4580413.<br>All repeats were identified using RepeatMasker: Smit, A.F.A. &<br>Green, P. (1996-1997)<br><a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a><br>location/Qualifiers  |
| FEATURES | 1. .120766   |
| source   | /organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/clone="MRPC.971_F_3"<br>/clone_lib="Peter de Jong/ human PAC library"<br>/map="17"<br>/chromosome="17"<br>1285. .1324<br>/rpt_family="(CGTG)n"<br>complement(1764. .1935)<br>/rpt_family="MER91A"<br>8156. .8350<br>/rpt_family="MIR"<br>11510. .11540<br>/rpt_family="(T)n"<br>11876. .11927<br>/rpt_family="L2"<br>complement(12664. .12966)<br>/rpt_family="AluSg"<br>complement(13019. .13043)<br>/rpt_family="MIR"<br>13044. .13334<br>/rpt_family="AluJo"<br>complement(13335. .13393)<br>/rpt_family="MIR"<br>13430. .13699<br>/rpt_family="AluSg"<br>complement(13704. .13872)<br>/rpt_family="MER91A"<br>13954. .13992<br>/rpt_family="MER65B"<br>complement(14240. .14426)<br>/rpt_family="MER5A"<br>15195. .15325<br>/rpt_family="MIR"<br>15946. .15960<br>/rpt_family="Charlie4"<br>complement(15961. .16254)<br>/rpt_family="AluSx"<br>16255. .16285<br>/rpt_family="Charlie4"<br>16286. .16588<br>/rpt_family="AluSg"<br>16589. .16661<br>/rpt_family="Charlie4"<br>complement(16696. .16900)<br>/rpt_family="MIR"<br>17327. .17631<br>/rpt_family="AluSg"<br>17635. .17928<br>/rpt_family="AluSc"<br>complement(17932. .17987)<br>/rpt_family="MIR"<br>complement(18019. .18188)<br>/rpt_family="MER91A"<br>18539. .18655 |

```

repeat_region /rpt_family="MIR"
complement(18859..19170)
repeat_region /rpt_family="AluY"
19182..19205
repeat_region /rpt_family="AT-rich"
19206..19505
repeat_region /rpt_family="AluSc"
complement(19218..19229)
unsure /note="Single-stranded terminator coverage."
repeat_region 19506..19524
/rpt_family="AT-rich"
repeat_region complement(19525..19837)
/rpt_family="AluSg"
19838..19858
repeat_region /rpt_family="AT-rich"
20517..20822
repeat_region /rpt_family="AluY"
complement(21146..21433)
repeat_region /rpt_family="AluSp"
complement(21532..21835)
/rpt_family="AluSx"
repeat_region 23272..23500
/rpt_family="AluSg"
repeat_region 23518..23827
/rpt_family="AluSg"
24479..24589
repeat_region /rpt_family="C-rich"
24984..25165
repeat_region /rpt_family="MIR"
25207..25471
repeat_region /rpt_family="AluUo"
25474..25560
repeat_region /rpt_family="(TA)n"
complement(25637..25933)
repeat_region /rpt_family="AluY"
complement(25940..26250)
/rpt_family="AluSx"
26573..26637
repeat_region /rpt_family="(CAG)n"
26800..26939
repeat_region /rpt_family="(CAG)n"
complement(29224..29299)
/rpt_family="MER33"
29300..29606
repeat_region /rpt_family="AluY"
complement(29607..29624)
/rpt_family="MER33"
29625..29908
repeat_region /rpt_family="AluSg"
complement(29909..29969)
/rpt_family="MER33"
repeat_region complement(29997..30045)
/rpt_family="AluUo"
complement(30046..30333)
/rpt_family="AluSx"
30222..30254
unsure /note="Single-stranded coverage."
repeat_region complement(30338..30515)
/rpt_family="AluY"
30338..30358
unsure /note="Possibly 21 T's, possibly 20 T's."
30366
repeat_region /note="Probably A, possibly G."
complement(30516..30828)
/rpt_family="AluSg"
repeat_region complement(30829..30964)
/rpt_family="AluY"
repeat_region complement(30965..31058)
/rpt_family="AluUo"
repeat_region complement(31059..31214)
/rpt_family="MER33"
31829..31896
repeat_region /rpt_family="(CAG)n"

```

```

repeat_region 31991..32280
/rpt_family="AluSg"
repeat_region 32281..32442
/rpt_family="AluSg/x"
repeat_region 32753..32794
/rpt_family="(GGA)n"
repeat_region 33138..33190
/rpt_family="(TG)n"
repeat_region 33285..33315
/rpt_family="(CCCCCG)n"
repeat_region 34350..34489
/rpt_family="AluUo"
repeat_region 34490..34784
/rpt_family="AluSx"
repeat_region 34785..34963
/rpt_family="AluUo"

```

Query Match 10.0%; Score 289; DB 40; Length 120766;  
 Best Local Similarity 98.3%; Pred. No. 7e-50; Mismatches 5; Indels 0; Gaps 0;  
 Matches 292; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1035 agatcctcacaagacagggagctgcccagcagccaccacccctgagagacag 1094
|||||
DB 32593 AGATCCTCACCAAGACAGGGAGCTGCCAGCAGCCACCACCCCTGAGAGACAG 32652
|||||
QY 1095 aggaagagctgacgagcagcagcagcagcagcagcagcagcagcagcagc 1154
|||||
DB 32653 AGGAGAGCTGACGAGGAGCGAGGAGGCTTCTGGGGAGGAGACCCCTGACCATGCCCC 32712
|||||
QY 1155 gggagagctccacagagatgtagacacacacaggaagacctgagagagagagag 1214
|||||
DB 32713 GGGAGGCTCCACAGAGAGTGTAGACACACAGGAAGACTGTGAGAGAGAGAGAGAG 32772
|||||
QY 1215 aggaatggagagagagagagagatgcatccagggttaagagcagagagagagag 1274
|||||
DB 32773 AGGATGGGAGAGAGAGAGAGATGTCATCAGCTTAAGAGACGAGAGGCGAGAGTGCG 32832
|||||
QY 1275 ctgaagagagggccgaccttgagagagcctggtgctgagatacaaaaaacgtctcag 1331
|||||
DB 32833 CTGAGAGAGGGCCCGACCTTGAGAGAGCGTGTGATGATACAAAAAATGACCCCG 32889
|||||

```

#### RESULT 11

```

AB018287 4238 bp mRNA PRI 16-JUN-1999
LOCUS AB018287
DEFINITION Homo sapiens mRNA for KIAA0744 protein, complete cds.
ACCESSION AB018287
VERSION AB018287.1 GI:3882208
KEYWORDS
SOURCE Homo sapiens adult male brain cDNA to mRNA, clone.lib:pbluescriptII
SK plus clone:hK04110.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

```

REFERENCE
AUTHORS Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL DNA Res. 5 (5), 277-286 (1998)
MEDLINE 99087487

```

```

REFERENCE
AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-1998) to the DBJ/EMBL/GenBank databases. Osamu
Ohara,Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kiserazu, Chiba 292-0812, Japan
(E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)

```

```

FEATURES
source 1..4238
/organism="Homo sapiens"

```

```

/db_xref="taxon:9606"
/clone="hK04110"
/clone_1lb="PBluescriptII SK plus"
/dev_stage="adult"
/sex="male"
/tissue_type="brain"
/gene="KIAA0744"
151..1923
/gene="KIAA0744"
151..1923
/codon_start=1
/product="KIAA0744 protein"
/protein_id="BAA34464.1"
/db_xref="GI:3882209"
/translation="MHSISSVDVKSSEVPVGLPISPLDRLTDLRMMPVDPVREK
OIOEELLIDOOOIOIKOLLIAFROKHEMLTROHQAQOEHLKELLAIKOQELLEK
EKLLEOOROEVEYERHREBQQLPPLRKDKGRRAVASTVYKLOEFLISKATKDY
PTGNKHSVRKPKLWYTAHHTSLDOSSPLSGTSPSYTYLPQADADKDPKRT
ASEPNLKVSRRLKQVAERSSPLRKDDNVTSRKMEVTESSVSSSGSGPS
SPNNGPTGAVTENETSVLPPTPHAEQMSQORLIHEDSNMLSLYTPSLPNTTGL
PAYSQLNNSNSLKEKOKETORLROGVPLPGYGSIPASSHPHYTLLEGKPNSSH
QALLOHLLKEOMROOKILYAGGVPLHPOSPLTKRISPIGRTGKLPHPRLNRO
SAPLPOSTLAQVITQOOHQDFLEKOKYQOOIHMKLISISIQLKQPSHLEABEE
LQGDQAMQEDRAPSGNSTRSSACVDTLGQGVAVKVEFPVSDDEDQIDQEMESG
EQAFMOOVYIGKDLAPGVTKVII"
BASE COUNT 1340 a 816 c 850 g 1232 t
ORIGIN
Query Match 7.78; Score 222.4; DB 9; Length 4238;
Best Local Similarity 55.7%; Pred. No. 4,4e-36;
Matches 602; Conservative 0; Mismatches 421; Indels 57; Gaps 7;
QY 55 cctccacagcagcccaaaagctgaggagccaccagctcttgagcagagttccc 114
DB 660 CCGCATCCCAAGCTGTGGTACACGCTGCCACACACATCATTTGATCAAGCTCTCC 719
QY 115 tccccagagcgccccctggagcagctccctcctacaactgcttgctggcccta 174
DB 720 ACCCTTAG-----TGGAAATCTCCATCTACACAGTACACATTACAGAGACACA 770
QY 175 cgcagcagcagcagactccccctcgcaaaagcctctgaacccaactgaagtgcg 234
DB 771 AGATGCAAGAGATGATTTCCCTTCGAAAAATCGCTTGAGCCCAACTTGAAGTGGC 830
QY 235 ttaagagctaaagagagtgctgtagcgagagcagctccctcctgctgtagcaga 294
DB 831 GTCCAGGTTAAACAGAAAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 890
QY 295 tggagactgataagcacttaagaagagagcgtgtagatcacaagtgccggcgctg 354
DB 891 TGGAAATGTTGTCACTTCATTCAGAAAGCGAATGTTGAGGTGACAGA----- 939
QY 355 ggcgtgctcgtgtgtaacagcgacccgctcgccgccagctctcccaaca---gctc 411
DB 939 ---ATCTCAGTCAAGTACAGTCTCTCCAGGCTGTGTCGCCAGTTCCACCAAACTGGG 995
QY 412 ccaagacacatcgctgaagatgcttactgctgtagtcccccaacatccccatgag-- 470
DB 996 AACTGGAAGGTACTGATAAATGAGACTTGGTTTGGCCCTTCACCCCTCATGCGCGAGA 1055
QY 470 -atgctccctcagcagcagccctcctctgagcagctcccccaacagttcagcctcta 528
DB 1056 AATGCTTTCACAGCAAGCATTTCAATTCATGAGATTCATTCATGAACTGCAATCTTTA 1115
QY 529 caagctctcctctctgcgaacatctccctagagctgtagcagcgaagctcattacca 588
DB 1116 TACCTCTCTCTTCTTGGCCCAACATTACTTGGGGCTTCCCGCACTG-----CC 1163
QY 589 ctacacctactgctcccccgaagcgtgtagacacagcagagcgagcgagcgccct 648
DB 1164 ATCCAGCTCAATGCTTGATTCATCAAGAAAGACAGAGTGTGAGAG----- 1216

```

```

QY 649 ccagctccctgagcagagtgtagcagctgacccgagcagttcagacatcctatccc 708
DB 1216 -CAGACGCTTAGCGAAGAGTGTCTCTCGCTGGGACAGATGAGAGCAGACATCCGGCATC 1274
QY 709 tgcgtcctctgctgagctgtagcagctgagggcgagcaggaagcccccagggcatgctcc 768
DB 1275 TTCCAGCCACCCCTCATGTTACTTTAGAGGGGAAAGCCACCAACAGCAGCACCAGGCTCT 1334
QY 769 gctgcagcatgtgctgtgctgtagcagcccgccagcagagcaccctcatgct----- 824
DB 1335 CCTCAGCATTTATTATTGAAAGAACAAATGCGACAGCAAAAGCTTTGTTAGTGGTGG 1394
QY 824 -gtgccactccagcgagctccaccatagtagcaggtgtagaagctgtggccaccagcagtcg 882
DB 1395 AGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAGAGAAATTTCACCTGGCATTTAG 1454
QY 883 gacgtgtagcagactcccgagcagctgagccctcagccagcacttagtctcagcgcgc 942
DB 1455 AGGTACCCACAAATGGCCCTCAGACACCCCTTAACCGAACCAGTCGACCTTTGCC 1514
QY 943 gcaagatcccccagcagctgtagcagctgtagcagcagcagcagcagcagcagcagcagc 1002
DB 1515 TCAGAGCAGCTGGCTGACGTGATTCATTCACAGCAACACCAACGATTTTGGAGAAACA 1574
QY 1003 gaagcagaagcagcagcagctacagctgtagcagatcctcaccagcagcagcagcagcagc 1062
DB 1575 GAAGCAATTACAGCAGCAGATCCACATGACCAAACTGCTTTCGAATCTATTGAACAAC 1634
QY 1063 cagcagccccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1122
DB 1635 GAACCAACAGCAGCAGTCACTTGAGGAAGCAGAGGAAGAGCTTCAGGGGAGCCAGGCGAT 1694

```

```

RESULT 12
AC023855/c
LOCUS
DEFINITION
Homo sapiens chromosome 17 clone RP11-546M21 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION
AC023855
VERSION
AC023855.1 GI:6997304
KEYWORDS
HMG; HMGs_PHASED.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 55893)
REFERENCE
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-546M21
Unpublished
2 (bases 1 to 55893)
REFERENCE
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barina,N., Bede,F., Boguslavsky,L.,
Bouckgalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Riley,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lenockzy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGuirk,A., McKernan,K., McPeckers,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,Y., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Telamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
TITLE
Direct Submission
JOURNAL
Submitted (18-FEB-2000) Whitehead Institute/MIT Center for Genome

```

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L6242  
Center clone name: 546\_M\_21  
-----  
\* NOTE: This record contains 72 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
1  
\* 771: contig of 771 bp in length  
\*  
\* 772 1555: contig of 784 bp in length  
\*  
\* 1556 2341: contig of 786 bp in length  
\*  
\* 2342 3121: contig of 780 bp in length  
\*  
\* 3122 3922: contig of 801 bp in length  
\*  
\* 3923 4709: contig of 787 bp in length  
\*  
\* 4710 5472: contig of 763 bp in length  
\*  
\* 5473 6247: contig of 775 bp in length  
\*  
\* 6248 6936: contig of 689 bp in length  
\*  
\* 6937 7707: contig of 771 bp in length  
\*  
\* 7708 8494: contig of 787 bp in length  
\*  
\* 8495 9267: contig of 773 bp in length  
\*  
\* 9268 10035: contig of 768 bp in length  
\*  
\* 10036 10811: contig of 776 bp in length  
\*  
\* 10812 11565: contig of 754 bp in length  
\*  
\* 11566 12332: contig of 767 bp in length  
\*  
\* 12333 13050: contig of 718 bp in length  
\*  
\* 13051 13838: contig of 788 bp in length  
\*  
\* 13839 14633: contig of 795 bp in length  
\*  
\* 14634 15403: contig of 770 bp in length  
\*  
\* 15404 16188: contig of 785 bp in length  
\*  
\* 16189 16964: contig of 776 bp in length  
\*  
\* 16965 17781: contig of 817 bp in length  
\*  
\* 17782 18561: contig of 780 bp in length  
\*  
\* 18562 19335: contig of 774 bp in length

\* 19336 20112: contig of 777 bp in length  
\* gap of unknown length  
\*  
\* 20113 20888: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 20889 21651: contig of 763 bp in length  
\* gap of unknown length  
\*  
\* 21652 22419: contig of 768 bp in length  
\* gap of unknown length  
\*  
\* 22420 23208: contig of 789 bp in length  
\* gap of unknown length  
\*  
\* 23209 23992: contig of 784 bp in length  
\* gap of unknown length  
\*  
\* 23993 24768: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 24769 25539: contig of 771 bp in length  
\* gap of unknown length  
\*  
\* 25540 26302: contig of 763 bp in length  
\* gap of unknown length  
\*  
\* 26303 27084: contig of 782 bp in length  
\* gap of unknown length  
\*  
\* 27085 27876: contig of 792 bp in length  
\* gap of unknown length  
\*  
\* 27877 28636: contig of 760 bp in length  
\* gap of unknown length  
\*  
\* 28637 29419: contig of 783 bp in length  
\* gap of unknown length  
\*  
\* 29420 30186: contig of 767 bp in length  
\* gap of unknown length  
\*  
\* 30187 30964: contig of 778 bp in length  
\* gap of unknown length  
\*  
\* 30965 31748: contig of 784 bp in length  
\* gap of unknown length  
\*  
\* 31749 32537: contig of 789 bp in length  
\* gap of unknown length  
\*  
\* 32538 33313: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 33314 34089: contig of 776 bp in length  
\* gap of unknown length  
\*  
\* 34090 34882: contig of 793 bp in length  
\* gap of unknown length  
\*  
\* 34883 35680: contig of 798 bp in length  
\* gap of unknown length  
\*  
\* 35681 36450: contig of 770 bp in length  
\* gap of unknown length  
\*  
\* 36451 37224: contig of 774 bp in length  
\* gap of unknown length  
\*  
\* 37225 37997: contig of 773 bp in length  
\* gap of unknown length  
\*  
\* 37998 38778: contig of 781 bp in length  
\* gap of unknown length  
\*  
\* 38779 39553: contig of 775 bp in length  
\* gap of unknown length  
\*  
\* 39554 40319: contig of 766 bp in length  
\* gap of unknown length  
\*  
\* 40320 41098: contig of 779 bp in length  
\* gap of unknown length  
\*  
\* 41099 41890: contig of 792 bp in length  
\* gap of unknown length  
\*  
\* 41891 42681: contig of 791 bp in length  
\* gap of unknown length  
\*  
\* 42682 43440: contig of 759 bp in length  
\* gap of unknown length  
\*  
\* 43441 44208: contig of 768 bp in length  
\* gap of unknown length  
\*  
\* 44209 44971: contig of 763 bp in length  
\* gap of unknown length  
\*  
\* 44972 45757: contig of 786 bp in length  
\* gap of unknown length  
\*  
\* 45758 46538: contig of 781 bp in length  
\* gap of unknown length  
\*  
\* 46539 47304: contig of 766 bp in length  
\* gap of unknown length



Query Match 7.58; Score 217.2; DB 56; Length 55893;  
Best Local Similarity 94.18; Pred. No. 4,7e-35;  
Matches 225; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

47305 48089: contig of 785 bp in length  
48090 48845: gap of unknown length  
48846 49634: contig of 756 bp in length  
49634: gap of unknown length  
49635 50428: contig of 789 bp in length  
50428: gap of unknown length  
50429 51219: contig of 794 bp in length  
51219: gap of unknown length  
51220 52009: contig of 791 bp in length  
52009: gap of unknown length  
52010 52790: contig of 790 bp in length  
52790: gap of unknown length  
52791 53586: contig of 781 bp in length  
53586: gap of unknown length  
53586: contig of 796 bp in length  
53586: gap of unknown length

Query Match 7.58; Score 217.2; DB 56; Length 55893;  
Best Local Similarity 94.18; Pred. No. 4,7e-35;  
Matches 225; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

586 caactcacactcaactcctccctccgagctgtgacacagagagagccgagagagc 645  
50950 CACCTTCAACCCNAGGCTCTCCCTGAACTGTGACACAGCAGAGGAGGAGGAGG 50891

646 cctccagctccctgagagaggtgagcagctgacccgagagagagagagagagag 705  
50890 CCTCCAGTCCCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 50831

706 tctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 765  
50830 TCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 50771

766 cctctcagagatgt 824  
50770 CCTCTCAGAGATGT 50712

RESULT 13  
AC014460/c  
LOCUS  
DEFINITION  
AC014460 54589 bp DNA HMG 16-NOV-1999  
Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered  
pieces.  
AC014460  
AC014460.1 GI:6436875  
HTG: HTGS\_PHASE2.  
KEYWORDS  
SOURCE  
ORGANISM  
fruit fly  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscophora; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 54589)  
AUTHORS  
Adams, M. and Venter, J.C.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,  
Rockville, MD, USA  
COMMENT  
This sequence was identified as CDM:10209628 by the submitter.  
For further information on this sequence e-mail to fly@celera.com.  
\* NOTE: This is a 'working draft' sequence.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
FEATURES  
source  
1. 54589  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
BASE COUNT 15712 a 11768 c 11765 g 15344 t  
ORIGIN

Query Match 7.48; Score 214.6; DB 42; Length 54589;  
Best Local Similarity 52.68; Pred. No. 1.6e-34;  
Matches 678; Conservative 0; Mismatches 444; Indels 167; Gaps 3;

1471 gcaactctccaccacaagt 1530  
7075 GCACAAGGTCAACACCGGTTTGGCTTACGATCCGCTTATGCTGAAGCATTCGTATTTGG 7016

1531 cggagacacacagctgacccctgacatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1500  
7015 CGGAGCAATGCCGATCCGGAGCAGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6956

1591 gcaagagacagagctgt 1650  
6955 AAATGAACGATGT 6886

1651 tgaatccacacagctgt 1710  
6895 GGAGCTGAGACTGT 6843

1711 ccgagagagacagacagacagagagagagagagagagagagagagagagagagag 1770  
6843 ---CAGTGGCCAGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6788

1771 gctcctgt 1830  
6787 TTTGTGCTGCGGCGGCTTGGGT 6728

1831 cagtgctgt 1890  
6727 AACGCTGTCAGAGATGT 6668

1891 agagctcaagaatgt 1950  
6667 TGACCTCGGAGATGT 6608

1951 agcagtgagatgt 2010  
6607 GGCATGAGGCTTGT 6548

2011 gttagagctgt 2042  
6547 GATGCCGAGAGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6488

2042 -----tgag 2045

6487 ACCAATTTCCATATATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6428

2046 acattcacatgt 2105

6427 ATGTGATTCAGGAGATGT 6368

2106 tctctgcacatgt 2165

6367 TTTCCATACATGACAG 6308

2166 ttgtgtga----- 2174

6307 TGAGTTTCTCTGATATATTTGCTTATATCAGATGATTAATTAATTAATTAATTAATTAAT 6248

2174 -----gagacagagctgt 2215

6247 GTTTTATATGTCGGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6188

2216 ggt 2275

6187 GCACTTATATCCGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6128

2276 ccattgt 2335

6127 CCCATCGCGGAGAGCTTATATCCGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 6068

2336 gaaggaatctgt 2395

6067 ACCGCGCATCCGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6008



|    |      |  |      |       |        |                             |
|----|------|--|------|-------|--------|-----------------------------|
| Qy | 2396 | accaggaacgtatgacccctggcaaggccgggtgtgtctgctgagggagggccat      | 2455 | 831   | 1216:  | contig of 386 bp in length  |
|    |      |  |      | 1217  | 1296:  | gap of unknown length       |
| Db | 6007 | ACCCGCGAAGCTCTTACGTTGGCCACGGCAAGAGTGTGCTGACGAGGCGGCTAC       | 5948 | 1297  | 1959:  | contig of 663 bp in length  |
| Qy | 2456 | gaattacacccatctgtatgtaccttgaaagtgtgtctgtctgtcagtgtaaaag      | 2515 | 1960  | 2039:  | gap of unknown length       |
|    |      |  |      | 2040  | 2473:  | contig of 434 bp in length  |
| Db | 5947 | GATTCGGCCGCATCTGTGATTCCGCCAGAGAGTGTGTGGGCGCTGCTCGGCGATCCC    | 5888 | 2474  | 2553:  | gap of unknown length       |
|    |      |  |      | 2554  | 3068:  | contig of 515 bp in length  |
| Qy | 2516 | ctgcagcccttgatgagcagcagtccttgcaagcaaaagccaacatcaagcagtgccacg | 2575 | 3069  | 3148:  | gap of unknown length       |
|    |      |  |      | 3149  | 3680:  | contig of 532 bp in length  |
| Db | 5887 | GCTGCTCCCATTTGCCAGCCGCGAGCTGAGCGCCCTTGGCCAGATGCAATCAATACG    | 5828 | 3681  | 3760:  | gap of unknown length       |
|    |      |  |      | 3761  | 4064:  | contig of 304 bp in length  |
| Qy | 2576 | ctagaagaatcatcagagatccagagcaa                                | 2604 | 4065  | 4144:  | gap of unknown length       |
|    |      |  |      | 4145  | 4586:  | contig of 442 bp in length  |
| Db | 5827 | CTCCAGAGACGATGACCATACAGTTAAA                                 | 5799 | 4587  | 4666:  | gap of unknown length       |
|    |      |  |      | 4667  | 5114:  | contig of 448 bp in length  |
|    |      |  |      | 4677  | 5194:  | gap of unknown length       |
|    |      |  |      | 5195  | 5689:  | contig of 495 bp in length  |
|    |      |  |      | 5690  | 5769:  | gap of unknown length       |
|    |      |  |      | 5770  | 6330:  | contig of 561 bp in length  |
|    |      |  |      | 6331  | 6410:  | gap of unknown length       |
|    |      |  |      | 6411  | 7118:  | contig of 708 bp in length  |
|    |      |  |      | 7119  | 7198:  | gap of unknown length       |
|    |      |  |      | 7199  | 7811:  | contig of 613 bp in length  |
|    |      |  |      | 7812  | 7891:  | gap of unknown length       |
|    |      |  |      | 7892  | 8703:  | contig of 812 bp in length  |
|    |      |  |      | 8704  | 8783:  | gap of unknown length       |
|    |      |  |      | 8784  | 9264:  | contig of 481 bp in length  |
|    |      |  |      | 9265  | 9344:  | gap of unknown length       |
|    |      |  |      | 9345  | 9763:  | contig of 419 bp in length  |
|    |      |  |      | 9764  | 9843:  | gap of unknown length       |
|    |      |  |      | 9844  | 10763: | contig of 920 bp in length  |
|    |      |  |      | 10764 | 10843: | gap of unknown length       |
|    |      |  |      | 10844 | 11263: | contig of 420 bp in length  |
|    |      |  |      | 11264 | 11343: | gap of unknown length       |
|    |      |  |      | 11344 | 11810: | contig of 467 bp in length  |
|    |      |  |      | 11811 | 11890: | gap of unknown length       |
|    |      |  |      | 11891 | 12486: | contig of 596 bp in length  |
|    |      |  |      | 12487 | 12566: | gap of unknown length       |
|    |      |  |      | 12567 | 13756: | contig of 1190 bp in length |
|    |      |  |      | 13757 | 13836: | gap of unknown length       |
|    |      |  |      | 13837 | 14335: | contig of 499 bp in length  |
|    |      |  |      | 14336 | 14415: | gap of unknown length       |
|    |      |  |      | 14416 | 15145: | contig of 730 bp in length  |
|    |      |  |      | 15146 | 15225: | gap of unknown length       |
|    |      |  |      | 15226 | 15946: | contig of 721 bp in length  |
|    |      |  |      | 15947 | 16026: | gap of unknown length       |
|    |      |  |      | 16027 | 17048: | contig of 1022 bp in length |
|    |      |  |      | 17049 | 17128: | gap of unknown length       |
|    |      |  |      | 17129 | 18037: | contig of 909 bp in length  |
|    |      |  |      | 18038 | 18117: | gap of unknown length       |
|    |      |  |      | 18118 | 18671: | contig of 554 bp in length  |
|    |      |  |      | 18672 | 18751: | gap of unknown length       |
|    |      |  |      | 18752 | 19606: | contig of 855 bp in length  |
|    |      |  |      | 19607 | 19686: | gap of unknown length       |
|    |      |  |      | 19687 | 20302: | contig of 616 bp in length  |
|    |      |  |      | 20303 | 20382: | gap of unknown length       |
|    |      |  |      | 20383 | 21177: | contig of 795 bp in length  |
|    |      |  |      | 21178 | 21257: | gap of unknown length       |
|    |      |  |      | 21258 | 21650: | contig of 393 bp in length  |
|    |      |  |      | 21651 | 21730: | gap of unknown length       |
|    |      |  |      | 21731 | 22830: | contig of 1100 bp in length |
|    |      |  |      | 22831 | 22910: | gap of unknown length       |
|    |      |  |      | 22911 | 23669: | contig of 759 bp in length  |
|    |      |  |      | 23670 | 23749: | gap of unknown length       |
|    |      |  |      | 23750 | 24686: | contig of 937 bp in length  |
|    |      |  |      | 24687 | 24766: | gap of unknown length       |
|    |      |  |      | 24767 | 25980: | contig of 1214 bp in length |
|    |      |  |      | 25981 | 26060: | gap of unknown length       |
|    |      |  |      | 26061 | 27037: | contig of 977 bp in length  |
|    |      |  |      | 27038 | 27117: | gap of unknown length       |
|    |      |  |      | 27118 | 28678: | contig of 1561 bp in length |
|    |      |  |      | 28679 | 28758: | gap of unknown length       |
|    |      |  |      | 28759 | 30049: | contig of 1291 bp in length |

RESULT 14  
AC010993/c  
LOCUS  
DEFINITION AC010993 182015 bp DNA HTG 15-FEB-2000  
Drosophila melanogaster chromosome X clone BACR27005 (DB77) RPI-98  
27.0.5 map 11F-12A strain Y; cn bw sp, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 135 unordered pieces.  
AC010993  
VERSION AC010993.9 GI:6978352  
KEYWORDS HTG; HTGS\_PHASE1.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
REFERENCE 1 (bases 1 to 182015)  
Celisner, S.E., Adayani, A., Arcaina, T.T., Baxter, E., Blazey, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,  
Kearney, L., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,  
Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,  
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.  
Sequencing of Drosophila melanogaster  
Unpublished  
2 (bases 1 to 182015)  
Celisner, S.E., Adayani, A., Arcaina, T.T., Baxter, E., Blazey, R.G.,  
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,  
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,  
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,  
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,  
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,  
Pfeiffer, B., Poon, L., Sequaira, A., Sethi, H., Snit, E.,  
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and  
Rubin, G.M.  
Direct Submission  
Submitted (29-SEP-1999) Drosophila Genome Center, Lawrence Berkeley  
Laboratory, MS 64-121, Berkeley, CA 94720, USA  
On Feb 15, 2000 this sequence version replaced gi:6838492.  
For further information about this sequence, including its location  
and relationship to other sequences, please visit our sequence  
archive Web site (<http://www.fruitfly.org/sequence/>) or send email  
to [hdg@fruitfly.berkeley.edu](mailto:hdg@fruitfly.berkeley.edu). All contigs in this submission meet  
the following cutoffs: length >= 200 bases.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 135 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 456: contig of 456 bp in length  
\* 457 536: gap of unknown length  
\* 537 750: contig of 214 bp in length  
\* 751 830: gap of unknown length

30050 30129: gap of unknown length  
30130 31096: contig of 967 bp in length  
31097 31176: gap of unknown length  
31177 32243: contig of 1067 bp in length  
32244 32323: gap of unknown length  
32324 33240: contig of 917 bp in length  
33241 33320: gap of unknown length  
33321 34447: contig of 1127 bp in length  
34448 34527: gap of unknown length  
34528 36019: contig of 1492 bp in length  
36020 36099: gap of unknown length  
36100 36950: contig of 851 bp in length  
36951 37030: gap of unknown length  
37031 38110: contig of 1080 bp in length  
38111 38190: gap of unknown length  
38191 39924: contig of 1734 bp in length  
39925 40004: gap of unknown length  
40005 40878: contig of 874 bp in length  
40879 40958: gap of unknown length  
40959 41929: contig of 971 bp in length  
41930 42009: gap of unknown length  
42010 43002: contig of 993 bp in length  
43003 43082: gap of unknown length  
43083 45063: contig of 1981 bp in length  
45064 45143: gap of unknown length  
45144 46258: contig of 1115 bp in length  
46259 46338: gap of unknown length  
46339 47436: contig of 1098 bp in length  
47437 47516: gap of unknown length  
47517 48865: contig of 1349 bp in length  
48866 48945: gap of unknown length  
48946 51872: contig of 2927 bp in length  
51873 51952: gap of unknown length  
51953 53613: contig of 1661 bp in length  
53614 53693: gap of unknown length  
53694 55676: contig of 1983 bp in length  
55677 55756: gap of unknown length  
55757 58351: contig of 2595 bp in length  
58352 58431: gap of unknown length  
58432 62335: contig of 3904 bp in length  
62336 62416: gap of unknown length  
62416 63858: contig of 1444 bp in length  
63859 63939: gap of unknown length  
63940 66720: contig of 2781 bp in length  
66721 66800: gap of unknown length  
66801 70313: contig of 3513 bp in length  
70314 70393: gap of unknown length  
70394 72962: contig of 2569 bp in length  
72963 73042: gap of unknown length  
73043 75757: contig of 2715 bp in length  
75758 75837: gap of unknown length  
75838 78597: contig of 2760 bp in length  
78598 78677: gap of unknown length  
78678 80980: contig of 2303 bp in length  
80981 81060: gap of unknown length  
81061 83621: contig of 2561 bp in length  
83622 83701: gap of unknown length  
83702 87265: contig of 3564 bp in length  
87266 87345: gap of unknown length  
87346 90782: contig of 3437 bp in length  
90783 90862: gap of unknown length  
90863 94530: contig of 3668 bp in length  
94531 94610: gap of unknown length  
94611 97589: contig of 2979 bp in length  
97590 97669: gap of unknown length  
97670 102001: contig of 4332 bp in length  
102002 102081: gap of unknown length  
102082 105930: contig of 3849 bp in length  
105931 106010: gap of unknown length  
106011 111112: contig of 5102 bp in length  
111113 111192: gap of unknown length  
111193 114833: contig of 3641 bp in length  
114834 114913: gap of unknown length

Query Match 7.4%; Score 214.6; DB 55; Length 182015;  
Best Local Similarity 52.6%; Pred. No. 1.5e-34;  
Matches 678; Conservative 0; Mismatches 444; Indels 167; Gaps 3;  
Qy 1471 gcaacttcaacaagaatggtgtctacagacagtcatactgaagaacagcagtcag 1530  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85737 GCACAGGTACACCCCGGTGGCTTACGATCCGCTTGTGGAAGCATTCGTATTTG 85678  
Qy 1531 cgggaacaacagtcgacccctgaagcagtcgacccgagtcgaagcagtcgacccg 1590  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85677 CGGAGACATATCCACGATCCGAGACATGTGTGACTGCGAGACGCTGTGGCACGGTT 85618  
Qy 1591 gcagagacagcagtcctctgaagtcgagcagcagtcgagcagtcgagcagtcgag 1650  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85617 AAATGAACGATCTGTGGAAGCGTTGGATCGGCTGGCGCTGCACAGCGACACAGA 85558  
Qy 1651 tgagatccagacagtcgacatcgaataccacacccctctctataggacagtcacctaa 1710  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85557 GGAGCTGCAGACTGTGCACACCGAAGCGCATGCTCTTCGGTTCCGAAT----- 85505  
Qy 1711 ccgagcagaagctagacagaagaagtcgtcgtcccaacagcagaagaatgtagtctgt 1770  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85505 -----CAGTCCAGCTCACGACGCCCCAAGTTGAAACACGTTTCGGCAGCTTTGTGCG 85450  
Qy 1771 gctgcctgtgtgggcatcgaggtgagacagtcgacagtcgtgagatgagatcctcctc 1830  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85449 TTTGTGCGTGGCGGCGCTTGGTGGTGTGATCTGGATACACGATGAGATGACACCATACGCC 85390  
Qy 1831 cagtcgtgtgcagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcgag 1890  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85389 AACCGCTGCAGGATGCGACGCGGTTGTGTATGATTTGCACATGGAAGCGCCACAGG 85330  
Qy 1891 agagctcagaatgattgattgcatcatccgccccagagacacacagcagagatccac 1950  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85329 TGACCTGCGGATGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 85270  
Qy 1951 agccatggatctgctctctcctcctcctcctcctcctcctcctcctcctcctcctc 2010  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85269 GGCATGGGCTTTGTTCTTCTTCAATTCATTCATTCATTCATTCATTCATTCATTCATTC 85210  
Qy 2011 gctgaacgtggcgaagtcctcctcctcctcctcctcctcctcctcctcctcctc 2042  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85209 GATGCCGAGGTGGCGGCGCATTCGATCGATTCGATTCGATTCGATTCGATTCGATTCG 85150  
Qy 2042 -----tgag 2045  
Db 85149 ACCAATTTCCATATTTGACTAAATTTACTAAATTTACTAAATTTACTAAATTTACT 85090  
Qy 2046 acattcaacatggaatggaacccagcagcagtcctcctcctcctcctcctcctcctc 2105  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85089 ATGTCATCAGGCGGATGCGACACAGCAAGCATTCATCAAAAGTCCGACATTCATTCAT 85030  
Qy 2106 tctctctgcatcgatatagaacaagggaactcctcctcctcctcctcctcctcctcctc 2165  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 85029 TTTTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 84970  
Qy 2166 ttgtgtgga----- 2174  
Db 84969 TGAGTTCTCTGATATTTTCGCTTAATCCATGCTGTAATTAACATTTTGTGTTTTTT 84910  
Qy 2174 -----gacacagcgttggtgatacctggaacgtgatacagcagcagcagcagc 2215  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 84909 GTTTATATGTCGGGCTCGGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTCGTGTC 84850  
Qy 2216 ggtgtgagcccccacatggaagcgtgagtagcctacagcctcagcagcagtcgtgag 2275  
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 84849 GCACCTATTCGCGGCTGCGGCGGCGGAGATATTCGCTGATTCGCTGATTCGCTGAT 84790  
Qy 2276 cccattgcccacaggttctcctcctcctcctcctcctcctcctcctcctcctcctcct 2335  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
Db 84789 CCATTCGCGGCGGAGCTTAAATCCGACATTCGATTCGCTGATTCGCTGATTCGCTGATTC 84730

| BASE COUNT | ORIGIN  |
|------------|---|
| 395 a      | 394 c 309 g 287 t   |
| OY         | 2336 gaagaaactcgtctccctctggtgctgactctgtcacgcgcaga1g1tttggcaacttg 2395   |
| Db         | 84729 ACCGGCATTCCGGGACCGCTGGGTGGTACCAATGTTCTTCGGGCTGCTGGGTTCATG 84670   |
| OY         | 2336 accaagcagctgatacaccctggcaaggcg9ggtgctgctgacccctggaggagccat 2455  |
| Db         | 84669 ACCCGCAACTCCTTCAGTTGGCCCAACGGCAAAAGTGTGCTGTGGCCTTGAGAGGCGGCTAC 84610  |
| OY         | 2456 gacttgacccgcacatctgtgatagtcctctgaagcttg1ctcgtgctctgctaag1tgaag 2515  |
| Db         | 84609 GATGCGCCGCATCTGTGATATCCGCCACAGGAGTGTGTGCGGGCGCTGCTCGGCATGCC 84550   |
| OY         | 2516 ctgcagccctctggaatgaagcagcttgctgaagcaaaagcccaacataagcagctgcccag 2575  |
| Db         | 84549 GCTGCTCCATTCATGCCACAGCCGACGCTGAGGGCCGCCCTTGCCAGAAATGCATCAATACG 84490  |
| OY         | 2576 ctagaagaatcatcgatgcagaagcaaa 2604  |
| Db         | 84489 CTCGAGAAAGCATATGACATACAGGTAA 84461  |
| RESULT 15  |   |
| AF124924   | AF124924 1385 bp mRNA PR1 27-OCT-1999   |
| LOCUS      | Homo sapiens histone deacetylase HDAC7 (HDAC7) mRNA, partial cds.   |
| DEFINITION | AF124924  |
| ACCESSION  | AF124924.1 GI:6119845   |
| VERSION    |   |
| KEYWORDS   | human.  |
| SOURCE     | Homo sapiens  |
| ORGANISM   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |
| REFERENCE  | 1 (bases 1 to 1385)   |
| AUTHORS    | Wang,A.H., Bertos,N.R., Vezmar,M., Pelletier,N., Crosato,M., Heng,H.H., Th'ng,J., Han,J. and Yang,X.J.  |
| TITLE      | HDAC7, a human histone deacetylase related to yeast HD1, is a transcriptional corepressor   |
| JOURNAL    | Mol. Cell. Biol. 19 (11), 7816-7827 (1999)  |
| REFERENCE  | 2 (bases 1 to 1385)   |
| AUTHORS    | Yang X.-J.  |
| TITLE      | Direct Submission   |
| JOURNAL    | Submitted (01-FEB-1999) Molecular Oncology Group, Department of Medicine, McGill University, 687 Pine Ave. W., Montreal, QC H3A 1A1, Canada   |
| FEATURES   | Location/Qualifiers   |
| SOURCE     | 1..1385   |
|            | /organism="Homo sapiens"  |
|            | /db_xref="taxon:9606"   |
|            | /chromosome="7"   |
|            | /map="7p15-p21"   |
|            | /tissue-type="brain"  |
|            | /dev_stage="fetus"  |
|            | <1..>1385   |
| gene       | /gene="HDAC7"   |
|            | <1..>1385   |
| CDS        | /note="similar to HDAC4"  |
|            | /codon_start=3  |
|            | /product="histone deacetylase"  |
|            | /protein_id="AA04254.1"   |
|            | /db_xref="GI:6119846"   |
|            | /translation="1LKEOKLEOOROEVEHRRRROOLPPLGKXGRGRRAVASNEVVKOLQERLLSKSATKDTPTNGKNHSVRHKMTYTAHHHSLDQSSP1LSTGSPSKRYTEGAQADKDDPFLKRTASEPILKVRSLKOKVAERSSPLRKDGNAVTSFKKRMFEVVESSSSSSPGSPSPNNGPTGVTENETVLPPTPAEQMVQOURLIHEDSMNLILHYTSPSLPNTTGLPAVPSQLMNSLSKEOKCETQIRGVPLPGYGSIPASSSHVHTLEKPPNSHOALQHLLEKQROOKLIVAGGVPLHPOSPLATKRISPGTIRPRLRHRRLPRTQASLPQSTLAQIVTIOOHQOEKOKYOQOIHMMNP1LEPTIRALSVRQAPLAAYGMDGLEHRLVSRKHSSPAASVLPHPAMDRLDGSATG1ATPDLMLKPPNS" |

|    |                       |   |                   |        |                 |                    |
|----|-----------------------|---|-------------------|--------|-----------------|--------------------|
|    | Query Match           | 7.1%  | Score 203.8       | DB 40: | Length 1385;    |                    |
|    | Best Local Similarity | 56.9%   | Pred. No.3.2e-32: |        |                 |                    |
|    | Matches               | 567;  | Conservative      | 0;     | Mismatches 367; | Indels 63; Gaps 8; |
| OY | 55                    | cctccacagacagcccacaatgctgtggagagaccacatgcttctlttgacaggagtcccc     | 114               |        |                 |                    |
| Dd | 218                   | ccgcgatcccaagctcgtgtagacagggcgccocacacacatcatatttgatccaagctcttc   | 277               |        |                 |                    |
| OY | 115                   | tcccagaagcgggcccccttgggaagcgctccctcctctaacaactgtcctttgcttgacctta  | 174               |        |                 |                    |
| Dd | 278                   | AACCCCTTAG-----TGGAAATCCTCCATCCATCAAGTAACAATTTACCAGGAGCGACA       | 328               |        |                 |                    |
| OY | 175                   | cgacagctcgagacagcatlccccctccgcgaanaaacgacctgaaccaacttgaagtatcg    | 234               |        |                 |                    |
| Dd | 329                   | AGATGCAAAAGGATGATTTCGCCCTTCGAAAAACTGCTCTGAAGCCAACTTGAAAGTGCg      | 388               |        |                 |                    |
| OY | 235                   | ttcaaagctaaacaagaagtggtgtgagcgagaaagcacagtcctccctcgtcglcgcaagga   | 294               |        |                 |                    |
| Dd | 389                   | GTCcAGGTTTAAACAGAAAGTGGCAGAGAAGGAGAGAGCACCCCTTACTAGCGCGAAGGA      | 448               |        |                 |                    |
| OY | 295                   | tggagctgttattagcacctttaaaaagagctgtttagatccaaggtgcggcgccg          | 354               |        |                 |                    |
| Dd | 449                   | TGGAAATGTGTCACTTCAATTCAMAAACGAAATGTTTAAAGTGACA-----               | 495               |        |                 |                    |
| OY | 355                   | ggcgtcgcgcgtgtgtlaaacgcgaccccggtcgcgcgccacagctctcccaaa--gctc      | 411               |        |                 |                    |
| Dd | 495                   | -GAATCCTCAGTCACTAGACAGTGTTCACAGGCTGTGTGCCAGTTACCAAMAACAAATGGGCGC  | 553               |        |                 |                    |
| OY | 412                   | ccacagacacatcgtctgagaatgcttacttgctlcagtlccccacaatccccactgag--     | 470               |        |                 |                    |
| Dd | 554                   | AACTGGAGGTGTACTGTGAAATGAGACACTTCGGTTTTGGCCCCCTACCCCTCATGCGGACGA   | 613               |        |                 |                    |
| OY | 470                   | -atgtccctcagacacgagccctccctcctctgaaagcctccccacaagcttgaagctcta     | 528               |        |                 |                    |
| Dd | 614                   | AATGTTTCACAGCAGAACGCATTCTAATTCAATGAAATTCATGAACCTGTCTAAGCTTTA      | 673               |        |                 |                    |
| OY | 539                   | caagctcctctctctgcgccacaatctccctaggagctgcagggccacgycatctytaccaa    | 588               |        |                 |                    |
| Dd | 674                   | TACCTCTCTCTTCTTGGCCCAACATTAACCTTGGGGCTTCCGCGATG-----CC            | 721               |        |                 |                    |
| OY | 589                   | ctcaaacctcaactgctctccccogaagctgttgcacacagcagagagggccgagagcgccct   | 648               |        |                 |                    |
| Dd | 722                   | ATCCCAGGTCTAATGTTGGAATTCACCTCAAAAGAAAGAGAAAGTGTGAGCGACAGACGCT     | 781               |        |                 |                    |
| OY | 649                   | ccagtcacctgcgcgaggggtgcagcctgaacgcgcgaagctcatgagcacatcctctatcc    | 708               |        |                 |                    |
| Dd | 782                   | TA-----GGCAAGGTGTCTCTCTGCTGTGGCAAGTATGGAAGCAGACATCCGGCATTC        | 832               |        |                 |                    |
| OY | 709                   | tggctgctgtctgtgcgttgcacitgagagggcgacgysagccccacgysgalgtccct       | 768               |        |                 |                    |
| Dd | 833                   | TYCCAGCCACCCCTCATGTTACTTGTAGAGGGAAGGACCCCAACACAGCACCAAGGCTCT      | 892               |        |                 |                    |
| OY | 769                   | gctgagacatgctgcctgtgtgcagcagggcccgagagagagcacccctatgtc-----       | 824               |        |                 |                    |
| Dd | 893                   | CCTGCAGCATTTATTATTATGAAAGAAACAAATGCGAAGCAAAAGCTTCTTGTAGCTGTGGG    | 952               |        |                 |                    |
| OY | 824                   | -gtgcacactccacgysagctccccactagtgacgsgtgtaaacgttgcgtgcacacagatcg   | 882               |        |                 |                    |
| Dd | 953                   | AGTTCCCTTACATCTCTCAAGTCTCCCTTGGCAACAAAGAGAAATTTACACCTGGCATTTAG    | 1012              |        |                 |                    |
| OY | 883                   | gaaggttaagcaagctcccgcgagcatcgcccttgagccgcgcacatcaagtcctcaecgctgc  | 942               |        |                 |                    |
| Dd | 1013                  | AGGTCGCCACAATAATGGCCCCGTCAACAGACCCCTGAAACCGAAACCCAGTGTGACACTTTGCC | 1072              |        |                 |                    |
| OY | 943                   | gcagagttccccgcgcctcgacgcagctgtcatactgacaacagacacgcagctttcctga     | 1002              |        |                 |                    |
| Dd | 1073                  | TCAGAGCACGTTGGCT-----CACTGTGTCTTTTAAACAGCAACCCAGCAATTTCTTGGA      | 1126              |        |                 |                    |
| OY | 1003                  | gaagcagaagcacgacagctacagcttgcggcaagatc                            | 1039              |        |                 |                    |
| Dd | 1127                  | GAAGCAGAAGCAATACCGACAGCAAGATTCACATGTGAAC                          | 1163              |        |                 |                    |

Search completed: May 20, 2000, 00:40:36  
Job time: 13040 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 20, 2000, 00:40:36 ; Search time 6491.04 Seconds

(without alignments)  
-194.527 Million cell updates/sec

Title: US-09-502-945-3

Perfect score: 1298

Sequence: 1 ggcctgctgaatgactgcga.....gaagaagaagaagaataa 1298

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0  
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_bal:\*  
2: gb\_baz:\*  
3: gb\_baz:\*  
4: gb\_ov:\*  
5: gb\_pat:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pr1:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_sts:\*  
14: gb\_sy:\*  
15: gb\_un:\*  
16: gb\_vl:\*  
17: em\_fun:\*  
18: em\_hum1:\*  
19: em\_hum2:\*  
20: em\_in:\*  
21: em\_om:\*  
22: em\_or:\*  
23: em\_ov:\*  
24: em\_pat:\*  
25: em\_ph:\*  
26: em\_pl:\*  
27: em\_ro:\*  
28: em\_sts:\*  
29: em\_sy:\*  
30: em\_un:\*  
31: em\_vl:\*  
32: gb\_htg1:\*  
33: gb\_htg2:\*  
34: gb\_in1:\*  
35: gb\_in2:\*  
36: em\_bal:\*  
37: em\_baz:\*  
38: em\_hum3:\*  
39: em\_hum4:\*  
40: gb\_pr4:\*  
41: gb\_htg3:\*  
42: gb\_htg4:\*  
43: gb\_htg5:\*  
44: gb\_htg6:\*

45: gb\_htg7:\*  
46: em\_htg1:\*  
47: em\_htg2:\*  
48: em\_htg3:\*  
49: em\_hum5:\*  
50: gb\_pl3:\*  
51: gb\_pr5:\*  
52: gb\_htg8:\*  
53: gb\_htg9:\*  
54: gb\_htg10:\*  
55: gb\_htg11:\*  
56: gb\_htg12:\*  
57: gb\_htg13:\*  
58: gb\_htg14:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID         | Description          |
|------------|-------|-------------|--------|---------------|----------------------|
| 1          | 1288  | 99.2        | 1298   | 40 AF039694   | AF039694 Homo sapi   |
| 2          | 729   | 56.2        | 5585   | 9 D87455      | D87455 Human mRNA    |
| 3          | 729   | 56.2        | 208577 | 33 AL135082   | AL135082 Homo sapi   |
| 4          | 308.6 | 23.8        | 472    | 13 G22194     | G22194 human STS W   |
| 5          | 215.4 | 16.6        | 7218   | 5 I66494      | I66494 Sequence 14   |
| 6          | 190.8 | 14.7        | 131025 | 52 AC022324   | AC022324 Homo sapi   |
| 7          | 189.6 | 14.6        | 53156  | 44 AC021195   | AC021195 Homo sapi   |
| 8          | 182.8 | 14.1        | 129404 | 54 AC013349   | AC013349 Homo sapi   |
| 9          | 182.4 | 14.1        | 198664 | 44 AC010649   | AC010649 Homo sapi   |
| 10         | 181.4 | 14.0        | 155284 | 41 AC009524   | AC009524 Homo sapi   |
| 11         | 180   | 13.9        | 76897  | 42 AC016179   | AC016179 Homo sapi   |
| 12         | 179.8 | 13.9        | 40482  | 11 HSAC000361 | HSAC000361 Human cos |
| 13         | 179.8 | 13.9        | 40977  | 11 HSAC000362 | HSAC000362 Human cos |
| 14         | 178.4 | 13.8        | 43346  | 54 AC011556   | AC011556 Homo sapi   |
| 15         | 178.2 | 13.7        | 62015  | 55 AC023326   | AC023326 Homo sapi   |
| 16         | 176.8 | 13.6        | 180385 | 40 AC007461   | AC007461 Homo sapi   |
| 17         | 176.2 | 13.6        | 42141  | 34 CEY54G9A   | AL032648 Caenorhab   |
| 18         | 176.2 | 13.6        | 300197 | 32 CEY54G9A   | 298869 Caenorhabd    |
| 19         | 175.8 | 13.5        | 140350 | 43 AC015830   | AC015830 Homo sapi   |
| 20         | 175   | 13.5        | 227070 | 56 AC023888   | AC023888 Homo sapi   |
| 21         | 174.4 | 13.4        | 179285 | 32 CNS01DS6   | AL121656 Homo sapi   |
| 22         | 174   | 13.4        | 879    | 13 CNS01JRG   | AL147405 Anopheles   |
| 23         | 174   | 13.4        | 73455  | 45 AC021589   | AC021589 Homo sapi   |
| 24         | 173.2 | 13.3        | 73020  | 54 AC022851   | AC022851 Homo sapi   |
| 25         | 172.2 | 13.3        | 68746  | 44 AC021008   | AC021008 Mus muscu   |
| 26         | 172   | 13.3        | 169424 | 42 AC011961   | AC011961 Homo sapi   |
| 27         | 171.2 | 13.2        | 120007 | 11 AF064864   | AF064864 Homo sapi   |
| 28         | 171.2 | 13.2        | 120609 | 40 AF165176   | AF165176 Homo sapi   |
| 29         | 170.8 | 13.2        | 175374 | 45 AC019265   | AC019265 Homo sapi   |
| 30         | 169.6 | 13.1        | 59815  | 44 AC019096   | AC019096 Homo sapi   |
| 31         | 169.2 | 13.0        | 129149 | 11 HS10D11    | 298044 Human DNA S   |
| 32         | 168.2 | 13.0        | 175132 | 40 AC008134   | AC008134 Homo sapi   |
| 33         | 168.2 | 13.0        | 123132 | 52 AC013455   | AC013455 Homo sapi   |
| 34         | 168   | 12.9        | 162575 | 33 AC004086   | AC004086 Homo sapi   |
| 35         | 167.8 | 12.9        | 175630 | 56 AC023059   | AC023059 Mus muscu   |
| 36         | 166.6 | 12.8        | 93615  | 40 AC004950   | AC004950 Homo sapi   |
| 37         | 166.6 | 12.8        | 142257 | 43 AC017491   | AC017491 Drosophi    |
| 38         | 166   | 12.8        | 190007 | 55 AC013314   | AC013314 Homo sapi   |
| 39         | 166   | 12.8        | 190538 | 33 AL135926   | AL135926 Homo sapi   |
| 40         | 165.2 | 12.7        | 34731  | 11 HS075931   | U75931 Human Xp22    |
| 41         | 165.2 | 12.7        | 99235  | 43 AC017010   | AC017010 Homo sapi   |
| 42         | 164.2 | 12.7        | 127778 | 44 AC020735   | AC020735 Homo sapi   |
| 43         | 164   | 12.6        | 220657 | 52 AC007500   | AC007500 Homo sapi   |
| 44         | 163.8 | 12.6        | 241350 | 58 AC021663   | AC021663 Homo sapi   |
| 45         | 163.2 | 12.6        | 247535 | 42 AC012645   | AC012645 Homo sapi   |

ALIGNMENTS

RESULT 1  
AF039694 1298 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-16 mRNA, complete cds.  
DEFINITION AF039694  
VERSION AF039694.1 GI:3170187  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 1298)  
Scanlan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
Gordan,J.D., Tureci,O., Sahin,U., Pfrendenschub,M. and Old,L.J.  
Characterization of human colon cancer antigens recognized by  
autologous antbodies  
Int. J. Cancer 76 (5), 652-658 (1998)  
JOURNAL  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 1298)  
Scanlan,M.J.  
AUTHORS  
TITLE Direct Submission  
Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1..1298  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colorectal carcinoma"  
11..406  
/codon\_start=1  
/product="antigen NY-CO-16"  
/protein\_id="AAC18043.1"  
/db\_xref="GI:3170188"  
/translation="MTANRLAESLIALSHOEELADLPKDYLLSEDEGNDGERKHQ  
KLEFAISLDGKNRKLAESEASLKYSEFNVSSESGEKLVLADLLEPKTSSSLAT  
VKKRLSIVKSKTVEPLINKEIERIHRE"  
BASE COUNT 548 a 190 c 381 g 174 t 5 others  
ORIGIN

Query Match 99.2%; Score 1288; DB 40; Length 1298;  
Best Local Similarity 99.6%; Pred. No. 3.5e-210;  
Matches 1293; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ggcgcgcgaatgactgcgaacgcgcttcgagagagccttcgtcttgagccancagga 60  
|||||  
Db 1 GGCgcgcgaatgactgcgaacgcgcttcgagagagccttcgtcttgagccancagga 60  
QY 61 agaactagcgcgatttcgcaaaaagactacctcttgagtgagagtgagaatgagggagaca 120  
|||||  
Db 61 AGAactagcgcgatttcgcaaaaagactacctcttgagtgagagtgagaatgagggagaca 120  
QY 121 tgatcgagagagaaagcatcnaaagctcttgaaagcaatcagcttccttga tggaaaga 180  
|||||  
Db 121 TGATcgagagagagaaagcatcnaaagctcttgaaagcaatcagcttccttga tggaaaga 180  
QY 181 tagcgagaaatgctgctgaaaggtctgaggtcagtcctgaaggtgtcagaggttcaatgctag 240  
|||||  
Db 181 TAGCGagaaatgctgctgaaaggtctgaggtcagtcctgaaggtgtcagaggttcaatgctag 240  
QY 241 ttctgaagatcacagaaagaaagctgctccttcgacagatcgtcttgaagcctgttaaaacttc 300  
|||||  
Db 241 TTCTGAAGATCACAGAAAGAAAGCTGtgccttcgacagatcgtcttgaagcctgttaaaacttc 300  
QY 301 atctctcttgccactgtgaaaaagcaactgagtagagtcanaatcaaagaanaacagtgga 360  
|||||  
Db 301 ATCTCTCTTGCCCACTGTGAAAAAGCAACTGAGTAGAGTCAAAATCAAGAAGACAGTGA 360  
QY 361 gttaaccttgacaagaagagatggaaggtccacagagagaatgacatcaataaagc 420  
|||||  
Db 361 GTTACCTCTGAAACAAGAGAGATTGAAGCGATCCACAGAGAAATAGCATTCATTAACG 420

QY 421 cacaagtcctcccaaatggagccctgtcgtcctcgaagaacccggcaggcagagcagcttg 480  
|||||  
Db 421 CACAAGTCCTCTCCAAATGGAGACCCTGTGTCCTCTGAAAGACCCGACAGCAAGCTGG 480  
QY 481 ttctcccttgagaaagagagcagccatgtctcccatgtgaacatgtgtcagtgct 540  
|||||  
Db 481 TTTTTCCTCCCTGAGAAAGAGAGCAGCCATTGCTCCCATGAACATGTGCTCAGTGCT 540  
QY 541 ggaagcgcaagatcccttccttgagcaggaatcttcaacctcttcataagaacaagcagc 600  
|||||  
Db 541 GGAAGCGCAAGATCCCTCTGAGCAGGAATAATTTCACCTCTCCATAAAGAACAGAGC 600  
QY 601 cagtgcagaccccttactgacccctgtgaaagagccctcctcgcagccatgagcctag 660  
|||||  
Db 601 CAGTGACAGACCCCTTACTGACCCCTGTGAAAAAGCCCTCTCTCGAGCCATGAGCCTAG 660  
QY 661 aagagcacaagatgcagacgacagagcttcagagagctcgggctctgcagtcctactatg 720  
|||||  
Db 661 AAGAGCCAAAGATGCAGACGACAGAGCTTCAGAGGCTCGGGCTCTGCAGTCTACTATG 720  
QY 721 angccaaggtctcgagagagagaaatcmaagtttaaagtatcacaaagtcgtgaaga 780  
|||||  
Db 721 ANGCCAAAGCTCGAAGAGAGAAATAATCNAAGTTAAACTATCACAAAGTCTGTAGA 780  
QY 781 aaggaagggccaagaaagccctaaagatttgacagctgcggaaggttaatccagctg 840  
|||||  
Db 781 AAGGAAGGCCCAAGAAAGCCCTAAAGAGTTTGAGCAAGCTCGGAGAGTTAATCCACTG 840  
QY 841 ccgcactagaagaaacgaagaaagaaagaaagagagagagagaaagaaagaaacaaag 900  
|||||  
Db 841 CCGCACTAGAAAGACGAAGAAAGAGAAAGAGAAAGAGAGAGAGAGAAAGAAAGAAAG 900  
QY 901 agagaagaaagaaag 960  
|||||  
Db 901 AGAAGAAAGAAAGAAAG 960  
QY 961 gagaagaaag 1020  
|||||  
Db 961 GAGAAGAAAG 1020  
QY 1021 agaggaagaaataagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaa 1080  
|||||  
Db 1021 AGAGGAAGAAATAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1080  
QY 1081 aggaagaaag 1140  
|||||  
Db 1081 AGGAAG 1140  
QY 1141 aacnagaagaaag 1200  
|||||  
Db 1141 AACNAGAAAG 1200  
QY 1201 ggaagaaag 1260  
|||||  
Db 1201 GGAAGAAAG 1260  
QY 1261 agaaagtataagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaagaaag 1298  
|||||  
Db 1261 AGAAAGTATAAGAAAG 1298

RESULT 2  
DB7455 5885 bp mRNA PRI 10-JUL-1997  
LOCUS Human mRNA for KIAA0266 gene, complete cds.  
DEFINITION DB7455  
ACCESSION DB7455.1 GI:1665798  
VERSION DB7455.1  
KEYWORDS KIAA0266  
SOURCE Homo sapiens male bone marrow myeloblast cell line:KG-1 cDNA to  
mRNA, clone\_id:pb1uescript II SK clone:HA2755.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 585)

**TITLE** Direct Submission  
**JOURNAL** Submitted (27-AUG-1996) to the DDBJ/EMBL/GenBank databases. Nobuo





```

Db 14595 GAAAGGAAAGGCCAAGAAAGCCTTAAAGAGCTTTGAGCAGCTACAGAAAGTTATTCAC 14654
OY 839 tgcgcactagaagaacgaagaagaagagaaagagagagagagagaaagaagaacaa 898
Db 14655 TGTGGCACTGCAAGAAATGAAAAAATTGAAATGTCAGAAATGATGAAAGCAATGAGCCT 14714
OY 899 ggaagaagaaga 910
Db 14715 TAAGCACCACAAA 14726

RESULT 4
LOCUS G22194/c G22194 STS 31-MAY-1996
DEFINITION human STS WI-13764, sequence tagged site.
ACCESSION G22194
VERSION G22194.1 GI:1342520
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE human STS derived from sequences in dbEST and the UniGene collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euteria; Primates; Carnivora; Hominoidea; Hominidae; Homo.
REFERENCE 1 (bases 1 to 472)
AUTHORS Hudson,T
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
MAPPED STS
COMMENT Mapped STS
Unpublished (1995)

Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Cambridge Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: CAATGCCACAGTTGATTTAA
Primer B: CGGCTCTGCAGTCTACTA
STS size: 150
PCR profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 mM
Tag Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3

FEATURES
Source
Derived from dbEST (genbank accession R54592).
Location/Qualifiers
1..472
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="145.4 CR from top of Chr13 linkage group"
20..169
20..169
complement(150..169)
157 t 3 others
BASE COUNT 83 a 111 c 118 g 157 t 3 others
ORIGIN

```

| Query Match           | 23.8%   | Score 308.6       | DB 13    | Length 472  |
|-----------------------|---|-------------------|----------|-------------|
| Best Local Similarity | 86.7%   | Pred. No. 1.1e-43 |          |             |
| Matches 385           | Conservative 0  | Mismatches 52     | Indels 7 | Gaps 4      |
| OY 429                | ctctccaatgagaccctcgtcctctgaagacccgacagcagcagcagctgtttt--c     | 486               |          |             |
| Db 446                | CTCCCAATGSGAGACCCCTATCATCTGAAAGAACACAGCAGCAGCAGCAGCTGTTTCC    | 387               |          |             |
| OY 487                | cccttgagaaagagagccagcca--tctcctccattgaacatgty-ctcagtygtcgg    | 542               |          |             |
| Db 386                | CCCTGGGGAAGAGAGCAGCCACCCATTGCTCCCATTTGAACATGCGCTCAGTGGCTGG    | 327               |          |             |
| OY 543                | aaggaagaagact-ccccctggagcagaagaatttcaactccctccataagaacaagcgc  | 601               |          |             |
| Db 326                | AAGGAAAGAACTCCCTCTGGAGCAGAGAAATTTTTTAACCTCTTCATTAAGAAACGACGC  | 267               |          |             |
| OY 602                | atgacagaccccttactgacccctgltgaaagcctctcccgagccatgacctaga       | 661               |          |             |
| Db 266                | ACTGACAGATCCTTTACTGACCTCCATGTGAAGAAAGGCTCTCTCCAAAGCATAGGCTTGA | 207               |          |             |
| OY 662                | agaagcacaagatgacagcagcagcagcttcaagggctcgggctctgcagtcctactatga | 721               |          |             |
| Db 206                | AGAGCAAAAGATGCACCGCAGCAGACTTCAGAGGCTCGGCTCTGCACTCTACTATGA     | 147               |          |             |
| OY 722                | ngccaagctcgaagaagaagaanaatcmaaaqgttaaaagatcaaaagtgtgaagaa     | 781               |          |             |
| Db 146                | GGCCAAAGCTCCAAAGAGAAAGAAATCAAAAGTAAAGATGATCAAAAGTGTGAAGAA     | 87                |          |             |
| OY 782                | aggaagagccaaagaagccctaaagaagtttagcagctcgtcggaaggttaatccagctgc | 841               |          |             |
| Db 86                 | AGGAAGGCGCAAGAAAGCCTTTAAAGAGTTTAGCAGCAGCAGAGGTTAATCCACTCT     | 27                |          |             |
| OY 842                | cgcactagaagaacgaagaagaaga 865                                 |                   |          |             |
| Db 26                 | GGCATTGGAGAAATGGAAGAAAAA 3                                    |                   |          |             |
| RESULT 5              |   |                   |          |             |
| LOCUS                 | I66494  | 7218 bp           | DNA      | 28-DEC-1997 |
| DEFINITION            | Sequence 14 from patent US 5670367.                           |                   |          |             |
| ACCESSION             | I66494  |                   |          |             |
| VERSION               | I66494.1  | GI:2724471        |          |             |
| KEYWORDS              | Unknown.  |                   |          |             |
| SOURCE                | Unknown.  |                   |          |             |
| ORGANISM              | Unknown.  |                   |          |             |
| REFERENCE             | 1 (bases 1 to 7218)   |                   |          |             |
| AUTHORS               | Donner, F., Scheiflinger, F. and Falkner, F. Gunter.          |                   |          |             |
| TITLE                 | Recombinant fowlpox virus.                                    |                   |          |             |
| JOURNAL               | Patent: US 5670367-A 14 23-SEP-1997;                          |                   |          |             |
| FEATURES              | Location/Qualifiers   |                   |          |             |
|                       | 1..7218   |                   |          |             |
|                       | /organism="unknown"   |                   |          |             |
| BASE COUNT            | 1944 a 1491 c 1486 g 1929 t 368 others                        |                   |          |             |
| ORIGIN                |   |                   |          |             |
| Query Match           | 16.6%; Score 215.4; DB 5; Length 7218;                        |                   |          |             |
| Best Local Similarity | 5.1%; Pred. No. 5.6e-28;                                      |                   |          |             |
| Matches 21            | Conservative 354; Mismatches 33; Indels 0; Gaps 0;            |                   |          |             |
| OY 846                | ctagaagaacgaagaagaagaagaagagagagagaagaagaagaacaagaagaag       | 905               |          |             |
| Db 1466               | CAAGTAAGTTAAAGACATGAGAAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRR      | 1407              |          |             |
| OY 906                | aagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag     | 965               |          |             |
| Db 1406               | RR        | 1347              |          |             |
| OY 966                | agaaagagagaagaagaagaagaagaagaagaagaagaagaactaagaagaagaag      | 1025              |          |             |

[illegible]

|       |        |                            |
|-------|--------|----------------------------|
| 5138  | 5147:  | gap of unknown length      |
| 5148  | 5640:  | contig of 493 bp in length |
| 5641  | 5650:  | gap of unknown length      |
| 5651  | 6146:  | contig of 496 bp in length |
| 6147  | 6156:  | gap of unknown length      |
| 6157  | 6649:  | contig of 493 bp in length |
| 6650  | 6659:  | gap of unknown length      |
| 6660  | 7154:  | contig of 495 bp in length |
| 7155  | 7164:  | gap of unknown length      |
| 7165  | 7658:  | contig of 494 bp in length |
| 7659  | 7668:  | gap of unknown length      |
| 7669  | 8031:  | contig of 363 bp in length |
| 8032  | 8041:  | gap of unknown length      |
| 8042  | 8533:  | contig of 492 bp in length |
| 8534  | 8543:  | gap of unknown length      |
| 8544  | 8861:  | contig of 318 bp in length |
| 8862  | 8871:  | gap of unknown length      |
| 8872  | 9365:  | contig of 492 bp in length |
| 9364  | 9373:  | gap of unknown length      |
| 9374  | 9638:  | contig of 265 bp in length |
| 9639  | 9648:  | gap of unknown length      |
| 9649  | 10141: | contig of 493 bp in length |
| 10142 | 10151: | gap of unknown length      |
| 10152 | 10643: | contig of 492 bp in length |
| 10644 | 10653: | gap of unknown length      |
| 10654 | 11143: | contig of 492 bp in length |
| 11146 | 11155: | gap of unknown length      |
| 11156 | 11647: | contig of 492 bp in length |
| 11648 | 11657: | gap of unknown length      |
| 11658 | 12151: | contig of 494 bp in length |
| 12152 | 12161: | gap of unknown length      |
| 12162 | 12579: | contig of 418 bp in length |
| 12580 | 12589: | gap of unknown length      |
| 13084 | 13083: | contig of 494 bp in length |
| 13094 | 13095: | gap of unknown length      |
| 13094 | 13565: | contig of 472 bp in length |
| 13566 | 13573: | gap of unknown length      |
| 13576 | 14067: | contig of 492 bp in length |
| 14068 | 14077: | gap of unknown length      |
| 14078 | 14570: | contig of 493 bp in length |
| 14571 | 14580: | gap of unknown length      |
| 14581 | 14955: | contig of 372 bp in length |
| 14953 | 14962: | gap of unknown length      |
| 14963 | 15455: | contig of 493 bp in length |
| 15456 | 15465: | gap of unknown length      |
| 15466 | 15586: | contig of 121 bp in length |
| 15587 | 15596: | gap of unknown length      |
| 15597 | 16090: | contig of 494 bp in length |
| 16091 | 16100: | gap of unknown length      |
| 16101 | 16592: | contig of 492 bp in length |
| 16593 | 16602: | gap of unknown length      |
| 16603 | 16781: | contig of 179 bp in length |
| 16782 | 16791: | gap of unknown length      |
| 16792 | 17285: | contig of 494 bp in length |
| 17286 | 17295: | gap of unknown length      |
| 17296 | 17793: | contig of 498 bp in length |
| 17794 | 17803: | gap of unknown length      |
| 17804 | 18296: | contig of 493 bp in length |
| 18297 | 18306: | gap of unknown length      |
| 18307 | 18800: | contig of 494 bp in length |
| 18801 | 18810: | gap of unknown length      |
| 18811 | 19303: | contig of 493 bp in length |
| 19304 | 19313: | gap of unknown length      |
| 19314 | 19807: | contig of 494 bp in length |
| 19808 | 19817: | gap of unknown length      |
| 19818 | 19913: | contig of 98 bp in length  |
| 19916 | 19925: | gap of unknown length      |
| 19926 | 20191: | contig of 266 bp in length |
| 20192 | 20201: | gap of unknown length      |
| 20202 | 20227: | contig of 26 bp in length  |
| 20228 | 20237: | gap of unknown length      |
| 20238 | 20518: | contig of 281 bp in length |
| 20519 | 20528: | gap of unknown length      |

|       |        |                            |
|-------|--------|----------------------------|
| 20529 | 21208: | contig of 680 bp in length |
| 21209 | 21218: | gap of unknown length      |
| 21219 | 21212: | contig of 494 bp in length |
| 21213 | 21172: | gap of unknown length      |
| 21213 | 22218: | contig of 496 bp in length |
| 22219 | 22228: | gap of unknown length      |
| 22229 | 22722: | contig of 494 bp in length |
| 22723 | 22732: | gap of unknown length      |
| 22733 | 22992: | contig of 260 bp in length |
| 22993 | 23002: | gap of unknown length      |
| 23003 | 23446: | contig of 444 bp in length |
| 23447 | 23456: | gap of unknown length      |
| 23457 | 23949: | contig of 493 bp in length |
| 23950 | 23959: | contig of 493 bp in length |
| 23960 | 24432: | contig of 493 bp in length |
| 24453 | 24462: | gap of unknown length      |
| 24463 | 24956: | contig of 494 bp in length |
| 24957 | 24966: | gap of unknown length      |
| 24967 | 25458: | contig of 492 bp in length |
| 25459 | 25468: | gap of unknown length      |
| 25469 | 25960: | contig of 492 bp in length |
| 25961 | 25970: | gap of unknown length      |
| 25971 | 26453: | contig of 493 bp in length |
| 26464 | 26463: | gap of unknown length      |
| 26474 | 26967: | contig of 494 bp in length |
| 26968 | 26977: | gap of unknown length      |
| 26978 | 27472: | contig of 495 bp in length |
| 27473 | 27482: | gap of unknown length      |
| 27483 | 27975: | contig of 493 bp in length |
| 27976 | 27985: | gap of unknown length      |
| 27986 | 28352: | contig of 377 bp in length |
| 28363 | 28372: | gap of unknown length      |
| 28373 | 28432: | contig of 120 bp in length |
| 28493 | 28502: | gap of unknown length      |
| 28503 | 28994: | contig of 492 bp in length |
| 28995 | 29004: | gap of unknown length      |
| 29005 | 29678: | contig of 674 bp in length |
| 29679 | 29688: | gap of unknown length      |
| 29689 | 30073: | contig of 385 bp in length |
| 30074 | 30083: | gap of unknown length      |
| 30084 | 30576: | contig of 493 bp in length |
| 30577 | 30586: | gap of unknown length      |
| 30587 | 31078: | contig of 492 bp in length |
| 31079 | 31088: | gap of unknown length      |
| 31089 | 31593: | contig of 495 bp in length |
| 31584 | 31593: | gap of unknown length      |
| 31594 | 32088: | contig of 495 bp in length |
| 32089 | 32098: | gap of unknown length      |
| 32099 | 32593: | contig of 495 bp in length |
| 32594 | 32603: | gap of unknown length      |
| 32604 | 33098: | contig of 495 bp in length |
| 33099 | 33108: | gap of unknown length      |
| 33109 | 33604: | contig of 496 bp in length |
| 33605 | 33614: | gap of unknown length      |
| 33615 | 34111: | contig of 497 bp in length |
| 34112 | 34121: | gap of unknown length      |
| 34122 | 34614: | contig of 493 bp in length |
| 34615 | 34624: | gap of unknown length      |
| 34625 | 35118: | contig of 494 bp in length |
| 35119 | 35128: | gap of unknown length      |
| 35129 | 35619: | contig of 491 bp in length |
| 35620 | 35629: | gap of unknown length      |
| 35630 | 36122: | contig of 493 bp in length |
| 36123 | 36132: | gap of unknown length      |
| 36133 | 36561: | contig of 429 bp in length |
| 36562 | 36571: | gap of unknown length      |
| 36572 | 37062: | contig of 491 bp in length |
| 37063 | 37072: | gap of unknown length      |
| 37073 | 37567: | contig of 495 bp in length |
| 37568 | 37577: | gap of unknown length      |
| 37578 | 38072: | contig of 495 bp in length |
| 38073 | 38082: | gap of unknown length      |
| 38083 | 38576: | contig of 494 bp in length |

|   | *<br>38577<br>38587   | 38586: gap of unknown length<br>38883: contig of 297 bp in length |                     |
|---|---|---|---------------------|
| Query Match   | 14.7%;  | Score 190.8; DB 52;   | Length 131025;      |
| Best Local Similarity   | 59.6%;  | Pred. No. 6.1e-24;  |                     |
| Matches 336;  | Conservative 0;   | Mismatches 226;   | Indels 2; Gaps 1.   |
| OY 735  | aagagaagaataatcnaaaagttaaaagttacacaagtctgtgaagaagaaggccaag      | 794   |                     |
| Db 129823   | AAAAAAAAAGACGACAAAAAAGAAAGAAAAGAAAAAAGAAAAAAGAAAAA              | 129764  |                     |
| OY 795  | aaagccctaagagatttgaagcagctgcggaaagttatccgcgctcgccactagaaga      | 854   |                     |
| Db 129763   | AAAAAAAAAAATTAATAAACCCCCCCCCCAGGAAGAGAAAAAAGCAAAAGCAAAAGG       | 129704  |                     |
| OY 855  | cgaagaagaagyaagaagaaggaggaagaagaagaagaagaagaagaagaaga           | 914   |                     |
| Db 129703   | AAGAAAAACAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAA        | 129644  |                     |
| OY 915  | gaagygagaagaagaagaagaagaagygagaagaagaagaagaagaagaagaaga         | 974   |                     |
| Db 129643   | ACAAAGCGAAAAAAGAGCAAAAAAAGAAAAAGAAGGAA--AAGAAAAAGAAAGAAAAAGA    | 129586  |                     |
| OY 975  | gaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaataag        | 1034  |                     |
| Db 129585   | AAAAACAAAGAAAAAAGCAGAAAGGAGAAAAAGCAAAGAAAAAAGAAAGAAAGAAAAAG     | 129526  |                     |
| OY 1035   | aaggaagaagaagaagaagaagaagttnaagaagaagaagaagaagaagaagaagaaga     | 1094  |                     |
| Db 129525   | AAAAAAGGAGAAAAACAAAAAAAGAAAAATGTGAAGAGAAAGAAAGAAAAAAGACAGAAAGAA | 129466  |                     |
| OY 1095   | agaactnagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga      | 1154  |                     |
| Db 129465   | GAAAGGAAAAAGAAAAAGAAAGGAAAGGAGAGAGAGAAAAAAGAAAAAGAAAAAGGGG      | 129406  |                     |
| OY 1155   | agaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga       | 1214  |                     |
| Db 129405   | AAAAAAAAAGAAAGAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA     | 129346  |                     |
| OY 1215   | aaaagaagaagaagaagaagaagaagttagaagaagcggaagaagaagaagaagaatataaga | 1274  |                     |
| Db 129345   | AAAAAAGAAAGAAAAAGAAAGAAAAAAGAAAAAGAGAAAGAAAGAAAGAAAAAGAGC       | 129286  |                     |
| OY 1275   | ggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga       | 1298  |                     |
| Db 129285   | GAAAGGAAAAAAGAAAAAGAAAAAGAAAAA 129262                           |   |                     |
| RESULT 7  |   |   |                     |
| LOCUS AC021195/c  | AC021195  | 53156 bp  | DNA HTG 14-JAN-2000 |
| DEFINITION Homo sapiens clone RP11-780K2, LOW-PASS SEQUENCE SAMPLING. |   |   |                     |
| ACCESSION AC021195  |   |   |                     |
| VERSION AC021195.1  |   |   |                     |
| KEYWORDS GI:6693387   |   |   |                     |
| SOURCE HTG: HTGS_PHASeq.  |   |   |                     |
| ORGANISM human.   |   |   |                     |
|   | Homo sapiens  |   |                     |
|   | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;   |   |                     |
|   | Eutheria; Primates; Catarrhini; Hominiidae; Homo.               |   |                     |
| REFERENCE 1 (bases 1 to 53156)  |   |   |                     |
| AUTHORS Waterston,R.H.  |   |   |                     |
| TITLE The sequence of Homo sapiens clone                              |   |   |                     |
| JOURNAL Unpublished   |   |   |                     |
| REFERENCE 2 (bases 1 to 53156)  |   |   |                     |
| AUTHORS Waterston,R.H.  |   |   |                     |
| TITLE Direct Submision  |   |   |                     |
| JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington  |   |   |                     |
| UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,   |   |   |                     |
| MO 63108, USA   |   |   |                     |
| COMMENT   |   |   |                     |
|   | Center project name: NH0563P16.                                 |   |                     |
|   | * NOTE: This record contains 94 individual                      |   |                     |
|   | * sequencing reads that have not been assembled into            |   |                     |
|   | * contigs. Runs of N are used to separate the reads             |   |                     |

\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 424: contig of 424 bp in length  
\* 425 434: gap of unknown length  
\* 435 827: contig of 393 bp in length  
\* 828 837: gap of unknown length  
\* 838 1335: contig of 498 bp in length  
\* 1336 1345: gap of unknown length  
\* 1346 1858: contig of 513 bp in length  
\* 1859 1868: gap of unknown length  
\* 1869 2196: contig of 328 bp in length  
\* 2197 2206: gap of unknown length  
\* 2207 2606: contig of 400 bp in length  
\* 2607 2616: gap of unknown length  
\* 2617 3130: contig of 514 bp in length  
\* 3131 3140: gap of unknown length  
\* 3141 3396: contig of 286 bp in length  
\* 3397 3406: gap of unknown length  
\* 3407 3921: contig of 515 bp in length  
\* 3922 3931: gap of unknown length  
\* 3932 4446: contig of 515 bp in length  
\* 4447 4456: gap of unknown length  
\* 4457 4972: contig of 516 bp in length  
\* 4973 4982: gap of unknown length  
\* 4983 5633: contig of 651 bp in length  
\* 5634 5643: gap of unknown length  
\* 5644 6339: contig of 696 bp in length  
\* 6340 6349: gap of unknown length  
\* 6350 6864: contig of 515 bp in length  
\* 6865 6874: gap of unknown length  
\* 6875 7388: contig of 514 bp in length  
\* 7389 7398: gap of unknown length  
\* 7399 7912: contig of 514 bp in length  
\* 7913 7922: gap of unknown length  
\* 7923 8436: contig of 514 bp in length  
\* 8437 8446: gap of unknown length  
\* 8447 9087: contig of 641 bp in length  
\* 9088 9097: gap of unknown length  
\* 9098 9738: contig of 641 bp in length  
\* 9739 9748: gap of unknown length  
\* 9749 10262: contig of 514 bp in length  
\* 10263 10272: gap of unknown length  
\* 10273 10934: contig of 662 bp in length  
\* 10935 10944: gap of unknown length  
\* 10945 11577: contig of 633 bp in length  
\* 11578 11587: gap of unknown length  
\* 11588 12284: contig of 697 bp in length  
\* 12285 12294: gap of unknown length  
\* 12295 12983: contig of 689 bp in length  
\* 12984 12993: gap of unknown length  
\* 12994 13602: contig of 609 bp in length  
\* 13603 13612: gap of unknown length  
\* 13613 14240: contig of 628 bp in length  
\* 14241 14250: gap of unknown length  
\* 14251 14941: contig of 691 bp in length  
\* 14942 14951: gap of unknown length  
\* 14952 15648: contig of 697 bp in length  
\* 15649 15658: gap of unknown length  
\* 15659 16172: contig of 514 bp in length  
\* 16173 16182: gap of unknown length  
\* 16183 16696: contig of 514 bp in length  
\* 16697 16706: gap of unknown length  
\* 16707 17155: contig of 449 bp in length  
\* 17156 17165: gap of unknown length  
\* 17166 17597: contig of 432 bp in length  
\* 17598 17607: gap of unknown length  
\* 17608 18053: contig of 446 bp in length

\* 18054 18063: gap of unknown length  
\* 18064 18885: contig of 822 bp in length  
\* 18886 18895: gap of unknown length  
\* 18896 19597: contig of 702 bp in length  
\* 19598 19607: gap of unknown length  
\* 19608 20511: contig of 904 bp in length  
\* 20512 20521: gap of unknown length  
\* 20522 21216: contig of 695 bp in length  
\* 21217 21227: gap of unknown length  
\* 21228 21676 21685: gap of unknown length  
\* 21686 22359: contig of 674 bp in length  
\* 22360 22369: gap of unknown length  
\* 22370 22882: contig of 513 bp in length  
\* 22883 22892: gap of unknown length  
\* 22893 23381 23391: contig of 489 bp in length  
\* 23382 23392: gap of unknown length  
\* 23392 24315: contig of 924 bp in length  
\* 24316 24325: gap of unknown length  
\* 24326 24839: contig of 514 bp in length  
\* 24840 24849: gap of unknown length  
\* 24850 25364: contig of 515 bp in length  
\* 25365 25374: gap of unknown length  
\* 25375 26204: contig of 830 bp in length  
\* 26205 26214: gap of unknown length  
\* 26215 26977 26986: gap of unknown length  
\* 26977 27681 27681: contig of 685 bp in length  
\* 27682 28585: gap of unknown length  
\* 28586 28595: contig of 904 bp in length  
\* 28596 29262: gap of unknown length  
\* 29263 29272: contig of 667 bp in length  
\* 29273 29947: gap of unknown length  
\* 29948 29957: contig of 675 bp in length  
\* 29958 30471: gap of unknown length  
\* 30472 30481: contig of 514 bp in length  
\* 30482 31118: gap of unknown length  
\* 31119 31128: contig of 637 bp in length  
\* 31129 31642: gap of unknown length  
\* 31643 31652: contig of 514 bp in length  
\* 31653 32167: gap of unknown length  
\* 32168 32177: contig of 515 bp in length  
\* 32178 33388: gap of unknown length  
\* 32389 33398: contig of 211 bp in length  
\* 33399 33052: gap of unknown length  
\* 33053 33062: contig of 654 bp in length  
\* 33063 33604: gap of unknown length  
\* 33605 33614: contig of 542 bp in length  
\* 33615 34128: gap of unknown length  
\* 34129 34138: contig of 514 bp in length  
\* 34139 34650: gap of unknown length  
\* 34651 34660: contig of 512 bp in length  
\* 34661 35174: gap of unknown length  
\* 35175 35184: contig of 514 bp in length  
\* 35185 35698: gap of unknown length  
\* 35699 35708: contig of 514 bp in length  
\* 35709 36222: gap of unknown length  
\* 36223 36232: contig of 514 bp in length  
\* 36233 36746: gap of unknown length  
\* 36747 36756: contig of 514 bp in length  
\* 36757 37270: gap of unknown length  
\* 37271 37280: contig of 514 bp in length  
\* 37281 37771: gap of unknown length  
\* 37772 37781: contig of 491 bp in length  
\* 37782 38285: gap of unknown length  
\* 37782 38285: contig of 504 bp in length  
\* 38286 38295: gap of unknown length  
\* 38286 38810: contig of 515 bp in length  
\* 38296 38820: gap of unknown length  
\* 38821 38835: contig of 515 bp in length  
\* 38835 39345: gap of unknown length  
\* 39346 39807: contig of 462 bp in length  
\* 39808 39817: gap of unknown length

|   |       |        |                            |
|---|-------|--------|----------------------------|
| * | 39818 | 40332: | contig of 515 bp in length |
| * | 40033 | 40382: | gap of unknown length      |
| * | 40243 | 40986: | contig of 644 bp in length |
| * | 40387 | 40956: | gap of unknown length      |
| * | 40397 | 41510: | contig of 514 bp in length |
| * | 41511 | 41520: | gap of unknown length      |
| * | 41521 | 42034: | contig of 514 bp in length |
| * | 42035 | 42044: | gap of unknown length      |
| * | 42045 | 42556: | contig of 514 bp in length |
| * | 42559 | 42566: | gap of unknown length      |
| * | 42569 | 43082: | contig of 514 bp in length |
| * | 43083 | 43093: | gap of unknown length      |
| * | 43093 | 43607: | contig of 515 bp in length |
| * | 43608 | 43617: | gap of unknown length      |
| * | 43618 | 44132: | contig of 515 bp in length |
| * | 44133 | 44142: | gap of unknown length      |
| * | 44143 | 44657: | contig of 515 bp in length |
| * | 44658 | 44667: | gap of unknown length      |
| * | 44668 | 45079: | contig of 412 bp in length |
| * | 45080 | 45089: | gap of unknown length      |
| * | 45090 | 45786: | contig of 697 bp in length |
| * | 45787 | 45796: | gap of unknown length      |
| * | 45797 | 46311: | contig of 515 bp in length |
| * | 46312 | 46321: | gap of unknown length      |
| * | 46322 | 46833: | contig of 514 bp in length |
| * | 46834 | 46845: | gap of unknown length      |
| * | 46846 | 47359: | contig of 514 bp in length |
| * | 47360 | 47369: | gap of unknown length      |
| * | 47370 | 47841: | contig of 472 bp in length |
| * | 47842 | 47851: | gap of unknown length      |
| * | 47852 | 48365: | contig of 514 bp in length |

|                           |        |                    |           |               |
|---------------------------|--------|--------------------|-----------|---------------|
| Query Match               | 14.6%  | Score 189.6;       | DB 44;    | Length 53156; |
| Best Local Similarity     | 56.5%; | Pred. No. 1.1e-23; |           |               |
| Matches 336; Conservative | 0;     | Mismatches 259;    | Indels 0; | Gaps 0        |

|    |       |   |       |
|----|-------|---|-------|
| QY | 704   | tcgaagctccactatgngccaagctcgaagagagaagaacacnaagttaaafta            | 763   |
| Db | 21258 | TTTGCCCTCATTCCTCCCATCTGGGAAAGACNNNNNNNNNNAAAAAAAAAAAAA            | 21199 |
| QY | 764   | tcaacaagtcgtgaagaagaagcgcaagaagccctaagaagtttgagcaagtcgc           | 823   |
| Db | 21198 | AAGA              | 21139 |
| QY | 824   | gaaggftaatccagctgccgactagaagacqgaagaagsgaaagaagtagagag            | 883   |
| Db | 21138 | AAA             | 21079 |
| QY | 884   | aagagaagaagacaaagcggaagaaagaagaaaggggaagagagaagaagagga            | 943   |
| Db | 21078 | AAA             | 21019 |
| QY | 944   | agaagaaaaggaaggaaggaagaagaaggaaggaagaagaagaagaaggaag              | 1003  |
| Db | 21018 | AAA             | 20955 |
| QY | 1004  | aagaaactaagaagaaggaaggaagaaactaagaaggaagaagaagaagaagtaaaga        | 1063  |
| Db | 20958 | AAA             | 20899 |
| QY | 1064  | agaagaacgaaggaagaagaagaaggaagaaactagaagaagaagaagaggaagaag         | 1123  |
| Db | 20898 | AAA             | 20833 |
| QY | 1124  | aagaaagaagaataagagcaacmaagaagaaggaagaagaataagagaaggaagaaga        | 1183  |
| Db | 20838 | AAATTAATAAAGCAAAAAA | 20779 |
| QY | 1184  | aaagaagaagaaggaaggaagaaggaaggaagaagaagaagaagaagaagtag             | 1243  |
| Db | 20778 | AAA             | 20719 |
| QY | 1244  | aaagcggaagaagaagaagaactaagaaggaagaagaagaagaagaagaaga              | 1298  |

Db 20718 AA 20664

|            |                              |                             |             |
|------------|------------------------------|-----------------------------|-------------|
| RESULT     |                              | 8                           |             |
| AC013349/c |                              |                             |             |
| LOCUS      | AC013349                     | 129404 bp                   | DNA         |
| DEFINITION | Homo sapiens clone RP1-22K1, | HTG                         | 06-FEB-2000 |
| ACCSSION   | AC013349                     | LOW-PASS SEQUENCE SAMPLING. |             |
| VERSION    | AC013349.2                   | GI:6910730                  |             |
| KEYWORDS   | HTG; HTGS-PHASED.            |                             |             |
| SOURCE     | human.                       |                             |             |
| ORGANISM   | Homo sapiens                 |                             |             |

| REFERENCE   | TITLE   |
|---|---|
| 1 (bases 1 to 129404)   | Birren,B., Linton,L., Nussbaum,C. and Lander,E. |
| 2 (bases 1 to 129404)   | Homo sapiens, clone RP11-22K1 unpublished       |
| Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M. |   |

**TITLE** Direct Submission  
**JOURNAL** Submitted (06-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** On Feb 6, 2000 this sequence version replaced g1:6572406.

http://ftp.genome.washington.edu/RM/Repaskasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www.seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: I4134  
Center clone name: 22\_K\_1

NOTE: This record contains 151 individual sequencing reads that have not been assembled into contigs. Runs of N are used to separate the reads and the order in which they appear is completely arbitrary. Low-pass sequence sampling is useful for identifying clones that may be gene-rich and allow overlap relationships among clones to be deduced. However, it should not be assumed that this clone will be sequenced to completion. In the event that the record is updated, the accession number will be preserved.

|      |       |                            |
|------|-------|----------------------------|
| 1    | 930:  | contig of 930 bp in length |
| 931  | 1855: | contig of 925 bp in length |
| 1856 | 2759: | contig of 904 bp in length |
| 2760 | 3706: | contig of 947 bp in length |
| 3707 | 4566: | contig of 860 bp in length |
| 4567 | 5444: | contig of 878 bp in length |
|      |       | gap of unknown length      |









| REFERENCE   | AUTHORS  | TITLE  | JOURNAL  | REFERENCE                         |
|---|--|--|--|-----------------------------------|
| 1 (bases 1 to 155284)                                 | Rosen, L., Madan, A., Olin, S., Abbasi, N., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S., Madan, A., Ratcliffe, A., Shafter, T. and Hood, L. | Sequencing of human chromosome 15 D15S160-D15S115 region | Unpublished  | 20155                             |
| 2 (bases 1 to 155284)                                 | Rosen, L., Madan, A., Olin, S., Abbasi, N., Baradarani, L., Birditt, B., Bloom, S., Dors, M., Dickhoff, R., Harrison, G., James, R., Lasky, S., Madan, A., Ratcliffe, A., Shafter, T. and Hood, L. | Direct Submission  | Submitted (26-AUG-1999) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA | 20968                             |
| * NOTE: This record contains 192 individual           |  |  |  |                                   |
| * sequencing reads that have not been assembled into  |  |  |  |                                   |
| * contigs. Runs of N are used to separate the reads   |  |  |  |                                   |
| * and the order in which they appear is completely    |  |  |  |                                   |
| * arbitrary. Low-pass sequence sampling is useful for |  |  |  |                                   |
| * identifying clones that may be gene-rich and allows |  |  |  |                                   |
| * overlap relationships among clones to be deduced.   |  |  |  |                                   |
| * However, it should not be assumed that this clone   |  |  |  |                                   |
| * will be sequenced to completion. In the event that  |  |  |  |                                   |
| * the record is updated, the accession number will    |  |  |  |                                   |
| * be preserved.                                       |  |  |  |                                   |
| 1   | 794: contig of 794 bp in length  | gap of unknown length                                    | 20155  | gap of unknown length             |
| 795   | 1602: contig of 808 bp in length   | gap of unknown length                                    | 20968  | contig of 814 bp in length        |
| 1603  | 2437: contig of 835 bp in length   | gap of unknown length                                    | 20969  | gap of unknown length             |
| 2438  | 3257: contig of 820 bp in length   | gap of unknown length                                    | 21769  | gap of unknown length             |
| 3258  | 4053: contig of 796 bp in length   | gap of unknown length                                    | 22583  | contig of 814 bp in length        |
| 4054  | 4865: contig of 812 bp in length   | gap of unknown length                                    | 23385  | contig of 803 bp in length        |
| 4866  | 5659: contig of 794 bp in length   | gap of unknown length                                    | 23386  | gap of unknown length             |
| 5660  | 6468: contig of 809 bp in length   | gap of unknown length                                    | 24201  | 24200: contig of 815 bp in length |
| 6469  | 7263: contig of 795 bp in length   | gap of unknown length                                    | 25005  | 25005: contig of 805 bp in length |
| 7264  | 8077: contig of 814 bp in length   | gap of unknown length                                    | 25006  | 25821: contig of 816 bp in length |
| 8078  | 8872: contig of 795 bp in length   | gap of unknown length                                    | 25822  | 26627: contig of 806 bp in length |
| 8873  | 9684: contig of 812 bp in length   | gap of unknown length                                    | 26628  | 27436: contig of 809 bp in length |
| 9685  | 10479: contig of 795 bp in length  | gap of unknown length                                    | 27437  | 28229: contig of 793 bp in length |
| 10480   | 11292: contig of 813 bp in length  | gap of unknown length                                    | 28230  | 29054: contig of 825 bp in length |
| 11293   | 12090: contig of 798 bp in length  | gap of unknown length                                    | 29055  | 29852: contig of 798 bp in length |
| 12091   | 12904: contig of 814 bp in length  | gap of unknown length                                    | 29853  | 30662: contig of 810 bp in length |
| 12905   | 13695: contig of 791 bp in length  | gap of unknown length                                    | 30663  | 31461: contig of 799 bp in length |
| 13696   | 14508: contig of 813 bp in length  | gap of unknown length                                    | 31462  | 32271: contig of 810 bp in length |
| 14509   | 15312: contig of 804 bp in length  | gap of unknown length                                    | 32272  | 33067: contig of 796 bp in length |
| 15313   | 16128: contig of 816 bp in length  | gap of unknown length                                    | 33068  | 33880: contig of 813 bp in length |
| 16129   | 16926: contig of 798 bp in length  | gap of unknown length                                    | 33881  | 34674: contig of 794 bp in length |
| 16927   | 17739: contig of 813 bp in length  | gap of unknown length                                    | 34675  | 35488: contig of 814 bp in length |
| 17740   | 18538: contig of 799 bp in length  | gap of unknown length                                    | 35489  | 36286: contig of 798 bp in length |
| 18539   | 19353: contig of 815 bp in length  | gap of unknown length                                    | 36287  | 37097: contig of 811 bp in length |
| 19354   | 20154: contig of 801 bp in length  | gap of unknown length                                    | 37098  | 37896: contig of 799 bp in length |
|   |  |  | 37897  | 38707: contig of 811 bp in length |
|   |  |  | 38708  | 39517: contig of 810 bp in length |
|   |  |  | 39518  | 40333: contig of 816 bp in length |
|   |  |  | 40334  | 41127: contig of 794 bp in length |
|   |  |  | 41128  | 41941: contig of 814 bp in length |
|   |  |  | 41942  | 42737: contig of 796 bp in length |
|   |  |  | 42738  | 43549: contig of 812 bp in length |
|   |  |  | 43550  | 44347: contig of 798 bp in length |
|   |  |  | 44348  | 45160: contig of 813 bp in length |
|   |  |  | 45161  | 45959: contig of 799 bp in length |
|   |  |  | 45960  | 46773: contig of 814 bp in length |
|   |  |  | 46774  | 47575: contig of 802 bp in length |
|   |  |  | 47576  | 48384: contig of 809 bp in length |
|   |  |  | 48385  | 49187: contig of 803 bp in length |
|   |  |  |  | gap of unknown length             |

|   |       |        |                            |
|---|-------|--------|----------------------------|
| * | 49188 | 49996: | contig of 809 bp in length |
| * | 49997 | 50792: | contig of 796 bp in length |
| * | 50793 | 51607: | contig of 815 bp in length |
| * | 51608 | 52405: | contig of 798 bp in length |
| * | 52406 | 53218: | contig of 813 bp in length |
| * | 53219 | 54014: | contig of 796 bp in length |
| * | 54015 | 54824: | contig of 810 bp in length |
| * | 54825 | 55622: | contig of 798 bp in length |
| * | 55623 | 56432: | contig of 810 bp in length |
| * | 56433 | 57228: | contig of 796 bp in length |
| * | 57229 | 58038: | contig of 810 bp in length |
| * | 58039 | 58837: | contig of 799 bp in length |
| * | 58838 | 59648: | contig of 811 bp in length |
| * | 59649 | 60444: | contig of 796 bp in length |
| * | 60445 | 61257: | contig of 813 bp in length |
| * | 61258 | 62054: | contig of 797 bp in length |
| * | 62055 | 62868: | contig of 814 bp in length |
| * | 62869 | 63667: | contig of 799 bp in length |
| * | 63668 | 64481: | contig of 814 bp in length |
| * | 64482 | 65278: | contig of 797 bp in length |
| * | 65279 | 66092: | contig of 814 bp in length |
| * | 66093 | 66892: | contig of 800 bp in length |
| * |       |        | gap of unknown length      |

|    |                       |  |              |              |            |      |        |         |
|----|-----------------------|--|--------------|--------------|------------|------|--------|---------|
|    | Query Match           | 14.0%  | Score        | 181.4;       | DB         | 41;  | Length | 155286; |
|    | Best Local Similarity | 58.8%;   | Pred.        | No. 2.4e-22; |            |      |        |         |
|    | Matches               | 332;   | Conservative | 0;           | Mismatches | 229; | Indels | 4;      |
|    |                       |  |              |              |            |      | Gaps   | 2.      |
| Oy | 734                   | aagagagaagaataatcnaagttaaaagtataccaaagtcgtagaagaagaagccca        | 793          |              |            |      |        |         |
| Dd | 93056                 | AAGAAAAAAAAAAAAACGAAAAAAACAAAAAAAAGCGAAAAAAGCAAAAAAACAACAAAAAAA  | 93115        |              |            |      |        |         |
| Oy | 794                   | gaagacctaaagaagtgtgagcagctcggaagtgtaaataccaagctgcgcactagaaga     | 853          |              |            |      |        |         |
| Dd | 93116                 | GAAGGAGCGAAGAAGNGGGGAAAAAAAAAAAAAAAAAAAAAAAAAAGGGGNGGGGGGGGCG    | 93175        |              |            |      |        |         |
| Oy | 854                   | acgaagaaaagagagaaaagaagagagagaaaagaagaaccaaagagaaagaagaaga       | 913          |              |            |      |        |         |
| Dd | 93176                 | GGAAAGGAAAGGAAAGAAAAAACCAAAAAAAGAAAAAAGGAAAAAGAAAAAGAAAA         | 93235        |              |            |      |        |         |
| Oy | 914                   | agaagggagaaagagaaagaagaagagagaaaagaagaagaagaagaagaag             | 973          |              |            |      |        |         |
| Dd | 93236                 | AAAAA--AAAGCAAAAAGCAAAAAAGCAAAAAAANNAANAAGAGAAAAAACAACAAAAA      | 93293        |              |            |      |        |         |
| Oy | 974                   | agaagaaaaagagaaagaagagaaagaagaacctaagaagaagagagagaaatatna        | 1033         |              |            |      |        |         |
| Dd | 93294                 | AAAAAGAGAGAAAAAGAAAAAAGAAAAAAGAAAGAAAAAAGAAAGAAAAAAGA            | 93353        |              |            |      |        |         |
| Oy | 1034                  | gaaggaaagaaagaagaanaagctaaagaagaagaagaagaagaagaagaag             | 1093         |              |            |      |        |         |
| Dd | 93354                 | GAAGCAAAAGAAAGGAAAGAAAAAAGCAAAAAAGCA--AAGGAGCAAAAAAAGAAAAAGCGAAG | 93411        |              |            |      |        |         |

|    |       |   |       |
|----|-------|---|-------|
| QY | 1094  | aagaaacttaagaagaagaagagaggaagaaaagaagaagaataaagaaatlaaggacmagaagaag | 1155  |
| DB | 93412 | AAAGGAAAAAACAAGAAAAAAGAAGAAAAGAGAAAAGAAAAAGAAAAAGAAAAAGAAA          | 93471 |
| QY | 1154  | gagaaagaagaataagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga         | 1213  |
| DB | 93472 | AAAAAAAAAAAAAAAAGGGGAACAAAAAGCAAGCAAAAAAAAAAAAAAAAAGACGNNANNAACA    | 93531 |
| QY | 1214  | aaaaagaagaaaaaagaagaagaagttagaaagcggaaagaagaagaagaagaatlaaga        | 1273  |
| DB | 93532 | ACAGCAGAGGAAAGAGGAAAAAGCAGAAAAAAGAAAAAAGAAAAAAGAGAGCNA              | 93591 |
| QY | 1274  | aggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga       | 1298  |
| DB | 93592 | AAAAAAAAAAAAAAAAGACAGA  | 93616 |

|            |   |
|------------|---|
| RESULT     | 11  |
| AC016179/c |   |
| LOCUS      |   |
| DEFINITION | Homo sapiens chromosome 15 clone RP11-652K13 map 15, LOW-PASS<br>SEQUENCE SAMPLING. |
| ACCESSION  | AC016179  |
| VERSION    | AC016179.1 GI:6466615   |
| KEYWORDS   | HTG; HTGS_PHASE0.   |
| SOURCE     | human.  |
| ORGANISM   | Homo sapiens  |

| REFERENCE            | AUTHORS  | TITLE   | JOURNAL     | REFERENCE            |
|----------------------|--|---|-------------|----------------------|
| 1 (bases 1 to 76897) | Birten, B., Linton, L., Nusbaum, C. and Lander, E. | Homo sapiens chromosome 15, clone RP11-652K13 | Unpublished | 2 (bases 1 to 76897) |

**AUTHORS**  
Birtren, B., Linton, L., Nushbaum, C., Lander, E., Allen, N., Anderson, M., Baldwin, J., Barua, N., Beckerly, R., Boguslavsky, L., Boukhalter, B., Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Deatellano, K., Dewalt, K., Domino, M., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Grant, G., Haas, B., Hearford, A., Horton, L., Howland, J., C. Johnson, R., Jones, C., Kamm, L., Karatas, A., Klein, J., Lemockzy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N., McMan, P., McGuck, A., McKernan, K., McDonald, J., Meldrum, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P., Stange-Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J., Teisfay, S., Tirrell, H., Vassiliev, H., Vo, A., Wheeler, J., Wu, X., Wyman, D., Ye, W., Zimmer, A. and Zody, M.

**TITLE** Direct Submission  
**JOURNAL** Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 330 Charles Street, Cambridge, MA 02141, USA  
**COMMENT** All repeats were identified using RepeatMasker:  
 Smt, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
 ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIMR  
 Web site: <http://www.seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: I4221  
 Center clone name: 652\_K\_13

```

* NOTE: This record contains 84 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will

```

```
* be preserved.
* 1 909: contig of 909 bp in length
* * gap of unknown length
* * 910 1826: contig of 917 bp in length
* * gap of unknown length
* * 1827 2740: contig of 914 bp in length
* * gap of unknown length
* * 2741 3646: contig of 906 bp in length
* * gap of unknown length
* * 3647 4536: contig of 890 bp in length
* * gap of unknown length
* * 4537 5459: contig of 923 bp in length
* * gap of unknown length
* * 5460 6373: contig of 914 bp in length
* * gap of unknown length
* * 6374 7290: contig of 917 bp in length
* * gap of unknown length
* * 7291 8196: contig of 906 bp in length
* * gap of unknown length
* * 8197 9111: contig of 915 bp in length
* * gap of unknown length
* * 9112 10028: contig of 917 bp in length
* * gap of unknown length
* * 10029 10921: contig of 893 bp in length
* * gap of unknown length
* * 10922 11822: contig of 901 bp in length
* * gap of unknown length
* * 11823 12753: contig of 931 bp in length
* * gap of unknown length
* * 12754 13648: contig of 895 bp in length
* * gap of unknown length
* * 13649 14534: contig of 886 bp in length
* * gap of unknown length
* * 14535 15422: contig of 888 bp in length
* * gap of unknown length
* * 15423 16300: contig of 878 bp in length
* * gap of unknown length
* * 16301 17219: contig of 919 bp in length
* * gap of unknown length
* * 17220 18114: contig of 895 bp in length
* * gap of unknown length
* * 18115 19069: contig of 955 bp in length
* * gap of unknown length
* * 19070 19986: contig of 917 bp in length
* * gap of unknown length
* * 19987 20903: contig of 917 bp in length
* * gap of unknown length
* * 20904 21798: contig of 895 bp in length
* * gap of unknown length
* * 21799 22719: contig of 921 bp in length
* * gap of unknown length
* * 22720 23650: contig of 931 bp in length
* * gap of unknown length
* * 23651 24543: contig of 893 bp in length
* * gap of unknown length
* * 24544 25473: contig of 930 bp in length
* * gap of unknown length
* * 25474 26395: contig of 922 bp in length
* * gap of unknown length
* * 26396 27292: contig of 897 bp in length
* * gap of unknown length
* * 27293 28202: contig of 910 bp in length
* * gap of unknown length
* * 28203 29140: contig of 938 bp in length
* * gap of unknown length
* * 29141 30054: contig of 914 bp in length
* * gap of unknown length
* * 30055 30941: contig of 887 bp in length
* * gap of unknown length
* * 30942 31855: contig of 914 bp in length
* * gap of unknown length
* * 31856 32782: contig of 927 bp in length
* * gap of unknown length

32783 33693: contig of 911 bp in length
* * gap of unknown length
* * 33694 34605: contig of 912 bp in length
* * gap of unknown length
* * 34606 35505: contig of 900 bp in length
* * gap of unknown length
* * 35506 36428: contig of 923 bp in length
* * gap of unknown length
* * 36429 37360: contig of 932 bp in length
* * gap of unknown length
* * 37361 38287: contig of 927 bp in length
* * gap of unknown length
* * 38288 39213: contig of 926 bp in length
* * gap of unknown length
* * 39214 40105: contig of 892 bp in length
* * gap of unknown length
* * 40106 41013: contig of 908 bp in length
* * gap of unknown length
* * 41014 41926: contig of 913 bp in length
* * gap of unknown length
* * 41927 42835: contig of 909 bp in length
* * gap of unknown length
* * 42836 43682: contig of 847 bp in length
* * gap of unknown length
* * 43683 44579: contig of 897 bp in length
* * gap of unknown length
* * 44580 45496: contig of 917 bp in length
* * gap of unknown length
* * 45497 46427: contig of 931 bp in length
* * gap of unknown length
* * 46428 47372: contig of 945 bp in length
* * gap of unknown length
* * 47373 48278: contig of 906 bp in length
* * gap of unknown length
* * 48279 49189: contig of 911 bp in length
* * gap of unknown length
* * 49190 50116: contig of 927 bp in length
* * gap of unknown length
* * 50117 51029: contig of 913 bp in length
* * gap of unknown length
* * 51030 51949: contig of 920 bp in length
* * gap of unknown length
* * 51950 52845: contig of 896 bp in length
* * gap of unknown length
* * 52846 53756: contig of 911 bp in length
* * gap of unknown length
* * 53757 54647: contig of 891 bp in length
* * gap of unknown length
* * 54648 55577: contig of 930 bp in length
* * gap of unknown length
* * 55578 56493: contig of 916 bp in length
* * gap of unknown length
* * 56494 57403: contig of 910 bp in length
* * gap of unknown length
* * 57404 58332: contig of 929 bp in length
* * gap of unknown length
* * 58333 59271: contig of 939 bp in length
* * gap of unknown length
* * 59272 60299: contig of 1028 bp in length
* * gap of unknown length
* * 60300 61204: contig of 905 bp in length
* * gap of unknown length
* * 61205 62143: contig of 939 bp in length
* * gap of unknown length
* * 62144 63046: contig of 903 bp in length
* * gap of unknown length
* * 63047 63967: contig of 921 bp in length
* * gap of unknown length
* * 63968 64874: contig of 907 bp in length
* * gap of unknown length
* * 64875 65789: contig of 915 bp in length
* * gap of unknown length
* * 65790 66742: contig of 953 bp in length
```



[illegible]

|            |  |                 |
|------------|--|-----------------|
| RESULT     | 13   |                 |
| HSAC000362 |  |                 |
| LOCUS      | HSAC000362                                 | 40977 bp ms-DNA |
| DEFINITION | Human cosmid g136ab094, complete sequence. | PRI             |
|            |  | 12-MAR-1997     |

|          |            |            |
|----------|------------|------------|
| VERSION  | AC000362.1 | GI:1881561 |
| KEYWORDS | HTG.       |            |
| SOURCE   | human.     |            |

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 40977)  
Iadonato, S.P., Yu, J., Wong, G.K.-S., Magness, C.L., Green, E.D.,  
Green, P. and Olson, M.V.

| TITLE  | JOURNAL            | REFERENCE            |
|--|--------------------|----------------------|
| Large-scale MCD Mapping and Sequencing of Human Chromosome 7 | Unpublished (1996) | 2 (bases 1 to 40977) |

|         |  |
|---------|--|
| AUTHORS | Magness, C.L.  |
| TITLE   | Direct Submission  |
| JOURNAL | Submitted (12-MAR-1997) Human Genome Center, University of |

**COMMENT**  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

These values are not generally visible from the Genbank flat file format but are available as part of the ASN.1 file.

Digest Mapping. Comparison of the experimentally derived map fragments with sequence-predicted fragments is given below. Sequences below are available at [www.ncbi.nlm.nih.gov/1400/000131](http://www.ncbi.nlm.nih.gov/1400/000131).

alignments below a variable cutoff (approximately 400–600bp) are mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. The small ordered fragment sizes are consistent

predicted values, uniquely ordered fragment groups are separated by dashed lines.

|       |     |         |     |      |     |
|-------|-----|---------|-----|------|-----|
| Man   | Sec | Man     | Sec | Man  | Sec |
| ECORI |     | HindIII |     | NsiI |     |

| map     | seq     | map     | seq     | map     | seq     |
|---------|---------|---------|---------|---------|---------|
| 8150.50 | 8225.00 | 4342.30 | 4337.00 | 4100.36 | 4094.00 |

|         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|
| 2000.57 | 2026.00 | 2090.70 | 2115.00 | 966.40  | 966.00  |
| -----   | -----   | -----   | -----   | -----   | -----   |
| 2612.00 | 2632.00 | 1671.50 | 1671.00 | 2803.25 | 2825.00 |

|         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|
| 2022.00 | 2022.00 | 1071.00 | 1071.00 | 2003.00 | 2003.00 |
| 3470.00 | 3545.00 | 3134.29 | 3147.00 | 6234.00 | 6330.00 |

|         |         |         |         |         |         |
|---------|---------|---------|---------|---------|---------|
| 2340.67 | 2360.00 | 2193.17 | 2218.00 | 960.40  | 963.00  |
| 841.00  | 832.00  | 2826.50 | 2895.00 | 3781.50 | 3837.00 |
| 1113.86 | 1105.00 | 5334.00 | 5353.00 | 1366.57 | 1377.00 |
| 4286.86 | 4284.00 | 4629.29 | 4633.00 | 5241.00 | 5220.00 |
| 9368.00 | 9345.00 | 2475.20 | 2479.00 | 1756.57 | 1772.00 |
| 790.92  | 785.00  | 1840.83 | 1888.00 | 3981.00 | 3952.00 |
| 652.06  | 657.00  | 4596.86 | 4534.00 | 7703.00 | 7706.00 |

## FEATURES

### source

```
repeat_region
/clone_11b="E. Green Chromosome 7 YAC Resource"
13. .221
```

```
repeat_region      /rpt_family="ALU"  
complement(3556.  .3637)
```

```
repeat_region /rpt_family="MER3"  
complement(3599..3637)
```

```
repeat_region
/rpt_family="MER33"
complement(5525..5986)
cat 6011..59833
```

```
repeat_region
/IPC_family="MLI2B2"
7876.7983
/cont_family="MT3B2"
```

```
repeat_region
1PC_family= MLI2B2
8195. .8315
/rpt_family="MLT2B2"
```

```
repeat_region complement(9623. .9892)
/rpt_family="ALU"
```

```
repeat_region      complement(10461..10548)
                    /rpt_family="MIR"
```

```
repeat_region complement(11049, .11365)
/rpt_family="ALU"
```

```
repeat_region 132/4. .13563
/rpt_family="ALU"
14353
14358
```

```
repeat_region      16556    16632  
repeated_family="MIR"  
repeat_region      14203    .14338
```

```
repeat_region      /rpt_family="MLT1"
repeat_region      16852.16958
```

```
repeat_region      /rpt_family="L1"
complement(18538.  .18785)
```

```
repeat_region      /rpt_family="MER43"
20054.    .23894
```

```
repeat_region      /rpt_family="L1"
                    complement(33450..33737)
```

```
repeat_region      /rpl_family="ALU"  
                   38075.  .38265  
                   /rnt_family="MTB"
```

| COUNT   | /family- min          |
|---------|-----------------------|
| 14085 a | 7738 c 7162 g 11992 t |

|                     |        |                    |        |              |
|---------------------|--------|--------------------|--------|--------------|
| ary Match           | 13.98; | Score 179.8;       | DB 11; | Length 40977 |
| st Local Similarity | 67.1%; | Pred. No. 5.2e-22; |        |              |

Matches 267; Conservative 0; Mismatches 130; Indels 1;

[illegible][illegible]

|                           |        |                    |           |               |
|---------------------------|--------|--------------------|-----------|---------------|
| Query Match               | 13.98; | Score 179.8;       | DB 11;    | Length 40977; |
| Best Local Similarity     | 67.18; | Pred. No. 5.2e-22; |           |               |
| Matches 267; Conservative | 0;     | Mismatches 130;    | Indels 1; | Gaps 1        |

[illegible]











RESULT 1  
AF039699 2162 bp mRNA PRI 05-MAY-1999  
LOCUS Homo sapiens antigen NY-CO-37 (NY-CO-38) mRNA, complete cds.  
DEFINITION AF039699  
ACCESSION AF039699.1 GI:3170197  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2162)  
Scalian,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
Gordan,J.D., Tureci,O., Sahin,U., Pfreundschuh,M. and Old,L.J.  
Characterization of human colon cancer antigens recognized by  
autologous antibodies  
JOURNAL Int J Cancer 76 (5), 652-658 (1998)  
MEDLINE 98272252  
REFERENCE 2 (bases 1 to 2162)  
Scalian,M.J.  
AUTHORS Direct Submission  
TITLE Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
Avenue, New York, NY 10021, USA  
FEATURES  
source  
1. 2162  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="colon cancer metastasis to lung"  
1. 2162  
/gene="NY-CO-38"  
97. .1308  
/gene="NY-CO-38"  
/note="pdz domain protein; NY-CO-38 antigen splice  
variant"  
/codon\_start=1  
/product="antigen NY-CO-37"  
/protein\_id="AAC18048.1"  
/db\_xref="GI:3170198"  
/translation="MDRKVAREFRHKVDPLIENDAKDLYDLVLMYHOTMDVAVLV  
DKLYINERSRLPLDARPLPLKHQVRYDQUTRRRSKLEKVALDRLHPGLSLV  
RGLEFGGLFSLHLKGGQADSVLDQDETVRLNGSISCTHEVYINLRKTV  
SLKRVHGLIPVKSDEPLTMQYDVQVSESGVSGLSGSGNENNEKKVFLSV  
SRGLCSISSGPIQPGIETSHVKPGSLAEVGLIGDYLIVNGLVDSNLDRKLVN  
VKNRSRLSTIVAAAGRELFWTDRELRSLAEVGLIGDYLIVNGLVDSNLDRKLVN  
OEMERORKEIKAKAEENERYRKEMQIVVEEEKFKQMEWMDSKQILLPKTITA  
EYHVPILRKP"

BASE COUNT 528 a 573 c 622 g 439 t  
ORIGIN

Query Match 100.0%; Score 2162; DB 40; Length 2162;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2162; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cctggccggtcgcggtcgcggtcttccagctccttgccagcgccggcaccgcgaagaac 60  
DB 1 CCGGCGCCGCGCGCGCGCGCTTTCAGCTCCTGCAGCCGGGACCCGGAAGAAC 60

QY 61 gggctgtcaacagcagcagctggaacctgcccagcagatggaagaatggcccgagaa 120  
DB 61 GGGTCTGTCAACGACGACCTGTGACCTGCCAGCCATGAGCGAAGGTGGCCAGAA 120

QY 121 ttccgcataaagtgatattcttcgtatgtaaaatgatacgcagagaagactatctctatg 180  
DB 121 TTCGCGCATTAAGGTGATTTTCTGATTGAATGATGACAGAGACTATCTATGAT 180

QY 181 gtcgtgcaatgtaccacacagacatgacgctggcgtgtcgttggagaaacctgaaactg 240  
DB 181 GGTGCTGCAATGTACCAACAGACATGACGTGGCGCTGCTGTGGAGAACCTGGAAGCTG 240

QY 241 gtcataatgaaccacgacgctctgctctgtttgatagtcattcgcgcgtgatccactg 300

DB 241 gtcataatgaaccacgacgctctgctctgtttgatagtcattcgcgcgtgatccactg 300  
QY 301 aagcaccaggtggaatatgatcagctgacccccgcgtccaggaagcttgaaagagtg 360  
DB 301 AAGCACCAGGTGGAATATGATATGATGACGTGACCCCCGCGCTCCAGAGAACTGAAAGAGTG 360

QY 361 cgtctggaacctctgaccccccgaagcctcgagctgagtgcgtgctgagatct 420  
DB 361 CGTCTGACCGTCTGACCCCGGAGGCGCTGAGCTGAGTGTGCTGTGGCTGAGATT 420

QY 421 ggtctggtctcttcatctccaccatcaaaaggggttcaggcagacagcgtcgagctc 480  
DB 421 GGCTGTGGCTCTCTCAATCTCCACCTCATCAAGGCGGTGAGCAACAGCGTGGAGCTC 480

QY 481 caggtgaggaagacagctcgcagatcaatgatatccatctccctccctgataccatgag 540  
DB 481 CAGGTAGGAGGACAGATCTGTCGGATCAATGATATTCATCTCTCTGATCCCATGAG 540

QY 541 gaggtcatcaacctcatctgaaaccaaagaatactgtgtcatcaaaagtgaacacatcggc 600  
DB 541 GAGGTCAATCAACCTCATCTGAAACCAAGAAACTGTGTCCATCAAAAGTGAGACACATCGCG 600

QY 601 ctgattcccgctgaaaagcctctctgataggccctacttgagagatgttgatcagttt 660  
DB 601 CTGATCCCGGTGAAAAGCTCTCTGTATGAGCCCTCATCTTGAGATGTGATCATGATT 660

QY 661 gttctggaatctggggggtgcaagcagccttgagctcccccgaatctgggaaacaag 720  
DB 661 GTTCTGGAATCTGGGGGCTGCAAGCAGCCTTGAGCTCCCGCTGGAATCTGGGAAACAAG 720

QY 721 gagaagaaggtcttcatacagcctgtgtaagctcccgagagccttgctgcagcatlccagc 780  
DB 721 GAGAAAGAAGTCTTCATCAAGCCTGTGAGGCTCCGAGGCTTGAGTGCAGCATTTCCAGC 780

QY 781 ggcgccatccagaagccttgcatctttatcagcaatgtgaaacctggctccctgtctgct 840  
DB 781 GGCCCATCCAGAAAGCCTGTCATCTTATCAGCCAGTGAACCTGGCTCCCTGTCTGCT 840

QY 841 gaggctggaattggaataagggagacagatgtgtcgaagctaaatggcgctgactcttaac 900  
DB 841 GAGGTGGATTGGAGATTAAGGAGACAGATGTTGCAAGTAAATGGCTGCATCTTCTTAAC 900

QY 901 ctggaatcaaaaggaagctgtgtaaatgtctgtaaaaaatagccgaagcctgacatctccat 960  
DB 901 CTGGATCAACAAGAGGCTGTAAATGTGCTGAATAAATAGCCGACGCTGCACATCTCCATT 960

QY 961 gtaagctgcaagcttggccgggaagctgttcatgacagacccggagagcgttgcgaagcgcg 1020  
DB 961 GTAGCTGCAAGCTGGCCGGGAGCTGTTCATGACAGACCCGGAGCGGCTGCAAGGCGCGG 1020

QY 1021 cagcgtgaagctcagcggcaggaagccttcatgcaagaacggtcgatgagatccaac 1080  
DB 1021 CAGCGTGAAGCTCAGCGGCGAGAGAGCTTCTCATGACAGACCGGCTGCGCATGAGTCCAAC 1080

QY 1081 aagatctctccaggaagcaggaagatgtagcgycgaagaagaagaatgtgccagaag 1140  
DB 1081 AAGATCTCTCCAGGAGCAGCAGAGATGAGCGGCAAGAGAGAAATGATGCCAGAG 1140

QY 1141 gcaagcagaagaaatgagagataccggaaagagatggaacaaatgttgaagaaagaaag 1200  
DB 1141 GCAGCAGAGAAATGAGATACCGGAAGAGATGGAACAATGTGAAGAGAGAAAG 1200

QY 1201 aagtttaagaagaatggaagaagactgggctcaaaaggaacagctactcttgcttaa 1260  
DB 1201 AAGTTTAAGAACAATGGAAGAAGACTGGGCTCAAAAGAACAGCTACTCTTGCTTAA 1260

QY 1261 accatcactgtcaggtataccaccagtaccccttcgcaagcgaagtgtatcttcggaata 1320  
DB 1261 ACCATCACTGTGAGGTATACCAACAGTACCCCTTCCGAACCAAGTGTATTCGGAAATA 1320

QY 1321 tgaaggaagctttgacccctactatctatgttcaacccagagacagatcattgggaaagatgt 1380  
DB 1321 TGAAGGAAGCTTTGACCCCTACTATCTATGTTC AACCCAGAGACAGATCATTGGGAAAGATGT 1380

Db 1321 TGAGGAGGCTTTGACCCCTACTCTATGTTCACCCAGACAGATCATGGGAGGATGT 1380  
 QY 1381 cgggtctctacgcataaagaaggagatctcttagacctggccctggaaaggcgtgtgtga 1440  
 Db 1381 CCGGCTCTCAGCATCAAGAGGAGGATCTTAGACCTGGCCCTGGAAAGCGGTGTGA 1440  
 QY 1441 ctcccccttgggaagggtgtctcttctgtctgtatgatgcggggagctctggcgga 1500  
 Db 1441 CTCGCCCATTTGGGAGAGTGTGCTTTGCTGTGTATGAGGCGGAGCTGTGAGCGGA 1500  
 QY 1501 tggltgagcatgttgaagaaggagacagatcatgagcaatcaacgacaagatltgtacagacta 1560  
 Db 1501 TGGTGGCATGTTGAAAGAGGAGCGAGATCATGGCAATCAAGGCAAGATTGTGACAGACTA 1560  
 QY 1561 caacctgtctgaggtctgagctgctgctgcaagaaggcttgaaatcaggcgaggagctggat 1620  
 Db 1561 CACCTGTGGCTGAGGCTTACGCTGCCGTGCAAGAGGCTTGAAATCAGGCGGCGGACTGGAT 1620  
 QY 1621 cgaacctgtgtgtgctgctgctgcccccaaggagatgatgagatgagcttctgtct 1680  
 Db 1621 CGACCTGTGTGTGCTGCTGTGCCCCCAAGAGATGACGATGACCTGTGACTTTCTTGTCT 1680  
 QY 1681 gaagtcacaaaggagaaacccaatltcacgcgttagaagaagtgagctccggcccaact 1740  
 Db 1681 GAAGTCCAAAAGGGGAAACCAAAATTCACGCGTTAGAAACAGTGAGCTCCGGCCCACT 1740  
 QY 1741 cgtgaaacaaaggcctgagaccagcttgagagagagccacatgacacacacagatgga 1800  
 Db 1741 CGTGAAACAAAGGCTTGGGACACGCTTGAAGAGGCGCACATGACACACACAGATGCGCA 1800  
 QY 1801 tctcttgagacctgaatcatcatcacccagagatctcaaatctcccttgcccttgacacagag 1860  
 Db 1801 TCCTTGGGAGCTGATATTTATCACCCAGGAAATCTCAAACTCCCTTTGGCCCTGAACACAGGG 1860  
 QY 1861 ccagataaaggagacagctcgagccacttcttgaaaggccaatgttggagaaaggagagc 1920  
 Db 1861 CCAGATTAAGGAACAGCTCGGAGCCACTTTTGAAGGCCAATGTGGAGAAAGGAGAGC 1920  
 QY 1921 cagccgttggagagagatctcaagatccagactctcatcttctctcttgccagctg 1980  
 Db 1921 CAGCCGCTTGGGAGAAATCTCAAGGATTCAGACTCTTCTTCTGCGCCAGTG 1980  
 QY 1981 aattgtctctcccaagcttgggagacctctcttgaaagcccaataaagacccacttg 2040  
 Db 1981 AATTGTCTCTCCACACTTTGGGAGCTCTTCTTGAACCCCTAATAAAGACCCCACTGG 2040  
 QY 2041 agctctctctccatccctctctctgctgctctgctctaatgtctgagagatgtca 2100  
 Db 2041 AGCTCTCTCTCTCCATCCCTCTCTCTGCTGCTGCTTAATTGCTGCGCAGATTTGTCA 2100  
 QY 2101 ctccaaaccttaactctgagctcatataaataaacaagattatttccagcttaaaa 2160  
 Db 2101 CTCCAAACTTACTGTGAGCTCATTAATAAATAAACAAGATTATTTCACAGCTTAAAAA 2160  
 QY 2161 aa 2162  
 Db 2161 AA 2162

RESULT 2  
 AF039700 2236 bp mRNA PRI 05-MAY-1999  
 LOCUS AF039700 Homo sapiens antigen NY-CO-38 (NY-CO-38) mRNA, complete cds.  
 DEFINITION AF039700.1 GI:3170199  
 ACCESSION AF039700.1  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE  
 1 (bases 1 to 2236)  
 Scannan,M.J., Chen,Y.T., Williamson,B., Gure,A.O., Stockert,E.,  
 Gordan,J.D., Tureci,O., Sahin,U., Pfleundschuh,M. and Old,L.D.

TITLE Characterization of human colon cancer antigens recognized by  
 autologous antibodies  
 JOURNAL Int. J. Cancer 76 (5), 652-658 (1998)  
 MEDLINE 98272252  
 REFERENCE 2 (bases 1 to 2236)  
 AUTHORS Scannan,M.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (23-DEC-1997) Ludwig Institute for Cancer Research, New  
 York Branch at Memorial Sloan-Kettering Cancer Center, 1275 York  
 Avenue, New York, NY 10021, USA  
 location/Qualifiers  
 1..2236  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..2236  
 /gene="NY-CO-38"  
 97..2055  
 /gene="NY-CO-38"  
 /note="PDZ domain protein"  
 /product="antigen NY-CO-38"  
 /protein="1"  
 /db\_xref="GI:3170200"  
 /translation="MDRKVAERFRHVDPLIENDAEKDYLYDYLRMYHOTMDVAVLVG  
 DIKVIINPSRLPLDPAIRPLIPDKHOVYDQTPRPSRKIKFVRLDRAPGICISV  
 RGLERGCGLFLSHLIKGDADSVGLQVDEDLIRNGISISCTHEEYINLIRTKTY  
 SLKVRHIGLIPVKSPPDEPLIMQYDQFVSESGVGLSPNRRNKKRKYFTSLVG  
 SRLGCSISSGPIOKPFIISHKPGSLAEVGLIGDQIVENGDFSNLHKAEVN  
 VLKNSRLTISIVAAARELFTMRDLAEARORELORELMOKFLAMESNKILDEO  
 EMEKRRKEILQKAENBRYKREKQIVVEERKKRQMEEDMGSKEDLLPKTTA  
 EYHVPVLRKPKDQGVPEPLRPADIDGTEBQEDBPKEGPPYSMTPEQING  
 KVRLLRIRKESGLDALRGSGVDSPIGKVVSAVYERGAERHGGIYKQDENIHALGN  
 IYVDTLADADALQKAMNGDMDLVVAVCPKRYEDLFFLLSKGNDIHALGN  
 SELRPLVNTKPTSLERGHMTTHRWHPMDLNLSPNRLPLPALNMGQIRNSGHRFE  
 GCGCGKAASRLGEDLKDPSHSFPAAQ"

BASE COUNT 548 a 588 c 653 g 447 t  
 ORIGIN

Query Match 96.1%; Score 2078; DB 40; Length 2236;  
 Best local similarity 96.7%; Pred. No. 0;  
 Matches 2162; Conservative 0; Mismatches 0; Indels 74; Gaps 1;

QY 1 cctggccgggtctgagctcgagctcttccagctctctgagccgggacccgaagaac 60  
 Db 1 CCTGGCCCGGTCTGCGGGCTGCGGCTCTTCCAGCTCTCGGAGCGGCGCACCCGAAGGAAAC 60  
 QY 61 gggtcgtgcaagcagcagcttgagcctggccagcagcttgagcagcagcagcagcagc 120  
 Db 61 GGGTCGTGCAACGACGAGCTGGGCTGGCCAGCCATGGAACGGAAGTGGCCCGAGAA 120  
 QY 121 ttcggcataaaggltgatttcttgatgataatgatgcagaagaagactatctatgat 180  
 Db 121 TTCGGCATTAAGGTGGATTTTCTGATTTGAATAATGATGCAAGAAGACTATCTCTAATGAT 180  
 QY 181 gtgtcgtgaatgtacacacagacagacagcagcagcagcagcagcagcagcagcagc 240  
 Db 181 GTGTGCGGAATGTACACACAGACACATGACAGCTGGCCGCTCGTGGGAGACAGAAAGCTG 240  
 QY 241 gtatcatatgaaccagcagcctgtgctctgttgatgacatcttgagccatctggcagctaccat 300  
 Db 241 GTATCATATGAACCCAGCCTGTGCTCTGTGTGATGCCATTTGGCGGCTGATGCCACTG 300  
 QY 301 aagcacacagctggaataatgatcagctgaccccccggcgctccaggaagcctggaagagtg 360  
 Db 301 AAGCACACAGCTGGAATATGATCAGCTACCCCGCGGCTCCAGGAAGCTGAAGAGAGTG 360  
 QY 361 cgtctgagacgctctgaccccccagcagcagcagcagcagcagcagcagcagcagcagc 420  
 Db 361 CGTCTGAGACCGCTGTGACCCCGAAGGCGCTGGCCTGAGTGTGCGTGTGCGTGTGAGATT 420  
 QY 421 ggtctgtggtcttcatctccacactataaaggcggtgaagcagacagcgtcgagctc 480

Db 421 GGGTGGGGCTCTTCATCTCCACCTTCATCAAAAGGGGCTGAGGACAGACAGCGTGGGCTC 480  
 QY 481 caggttaggagacagatcgtcccgatcaatgatatccatctccctctgtaccatgag 540  
 Db 481 CAGGTAGGGAGACAGATCGTCCGGATCAATGGATATTCCATCTCTCTCTTACCCTATGAG 540  
 QY 541 gaaggtcaaacctctctgaaccagaanaactgtgtctcaatcaagttagacacatcgtgc 600  
 Db 541 GAGGTATCAACTCATTTGCAACCAAGAAACTGTGTCTCATCAAGATGAGACATCTGGC 600  
 QY 601 ctgattcccgtagaaagctctctgtatgagccctcaactgtgagatgtgtgacgttt 660  
 Db 601 CTGATCTCCGTGAAGAGCTCTCTGTGATGAGCCCTTCATGGCAGATGTGATGATGATTT 660  
 QY 661 gtctggaatctggggcgtgtcgagcaagcctgggtccctcggaatcggaatacaag 720  
 Db 661 GTCTCGGAATCTGGGGCGTGGCAGGACGCTGGGCTCCCTCGGAATCGGGAAGAACAG 720  
 QY 721 gagaagaaggtctctcaacgtgtgtaggtctccgagcctgtgtctgacatctccagc 780  
 Db 721 GAGAAAGAGGTCTTCATCAGCTGTGAGGCTCCGAGGCTTGCGTGGCATTTTCAGC 780  
 QY 781 gggcccatccagaagcctgtgcatcttatacagcatgtgaacctgtctccgtctgtc 840  
 Db 781 GGGCCCATCCAGAGCTGTGCATCTTTATCAGCCATGTGAACCTGCTCCCTGTCTGCT 840  
 QY 841 gaagtgtgatttggagataaggagaccagatgtctgaatgcatagcgtctgacatctctaac 900  
 Db 841 GAGGTGGGATTTGAGATTAGGGGACAGATTTGTGAATGCTAATGGCTGACATCTCTTAAC 900  
 QY 901 ctggtacacaaggagcctgttaaatgtgtctgtaaaatagccgacgtctgacatctccatt 960  
 Db 901 CTGTGATCAACAAGAGGCTGTAAATGTGTGTAATAATAGCCGACCTGACCATCTCATTT 960  
 QY 961 gtagctgcagcgtctggccggagcgtgtctcatagacagaccggagagcgtgtcgagagcgcg 1020  
 Db 961 GTAGCTGCAGCCTGGCCGGAGCTGTTCATGACAGCCGGAGGCGTGGCAGAGGGCGG 1020  
 QY 1021 caacgtgtagcgtgcagcggagcgtctctcatgcagaagcgtgtgcagatgagatcaac 1080  
 Db 1021 CAGCGTAGCTGCAGCGGACGAGAGCTTCTCATCAGAAAGCGCTGGCGATGGATGATCAAC 1080  
 QY 1081 aagatctctccagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1140  
 Db 1081 AAGATCTCTCCAGGACGAGCAGAGATGAGAGCGGCAAGAGGAAAGAAATTTGCCAGAG 1140  
 QY 1141 gtagcagagaaatagagatataccggaagagatagagacagctttagagagagagagag 1200  
 Db 1141 GCAGCAGAGAAATAGAGATACCGGAAGAGATGAGATGATGTAGAGAGAGAGAG 1200  
 QY 1201 aagtttaagaagaatgaatgagaaagcctgggtctcaaaagaacagctactctgtctaa 1260  
 Db 1201 AAGTTTAAGAAGCAATGGGAGAGAACTGGGCTTCAAGGAACAGCTACTCTTGCTTAA 1260  
 QY 1261 accatcaactgtgtgaggtacacacagtaaccctcgtgaagccaaagt----- 1307  
 Db 1261 ACCATCACTGTGAGGTACACCCAGTACCCCTTCGCAAGCAAAAGTATGATCAGGAGTGT 1307  
 QY 1307 ----- 1307  
 Db 1321 GAACCTGAGCTCGAGCCCGCAGATGACCTGATGAGGACGAGAGAGAGAGAGAGAGAG 1380  
 QY 1307 gattccggaataataggaagaagccttgaccctactctatgttcaaccacagaagcagatc 1366  
 Db 1381 GATTTCGGGAATATAGAGAGAGGCTTTGACCCCTACTCTATGTTCACCCCAAGAGCAGATC 1440  
 QY 1367 atgtgggaagaatgtccggtctctacgcatcaagaagaagagagatcctctagacctgtgcct 1426  
 Db 1441 ATGAGGGAAGATGTCCGGCTCTTACCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1427 gaagcggtgtgtgagcctcccatctgggaaggtgtgtcgttctgtcgtgtatagagcgagga 1486  
 Db 1501 GAAGCGGAGTGTGAGATCCCATTTGGGAAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1560

QY 1487 gctctgaagcgcagatgtgtgcatctgtgaaagggagacagatcatggaatcaacagcag 1546  
 Db 1561 GCTGTGAGAGGAGATGTGGCATTTGAAAGGGAGAGAGATCTATGGCAATCAAGGGCAAG 1620  
 QY 1547 attgtgacagactacacccctgtgtgagcgtgtgacgtgtccctgcagaagcctgtgaatcag 1606  
 Db 1621 ATTGTGACAGACTACACCCCTGAGCTGAGCGTGAAGCTGTCCCTGCAGAGGCTGTGAATCAG 1680  
 QY 1607 ggcgggagactgtagactgt 1666  
 Db 1681 GGGGGGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740  
 QY 1667 ctgacctctctgtgtagtccaaaggggaacaaatcaacgcgtctgaagaaacagctgag 1726  
 Db 1741 CTGACCTCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800  
 QY 1727 ctccggcccaactcgttgaacacaaagcctcggacagccttgaagagagccacatgaca 1786  
 Db 1801 CTCCGGCCCACTCGTGAACACAAAGCCTCGGACAGCCTTGAGAGAGGCGCATGACA 1860  
 QY 1787 cacaccagatgtagtctccttggagcctgaaatctatacccaagagatctcaactcccttg 1846  
 Db 1861 CACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920  
 QY 1847 ggcctgaacagagccagataaagaaacagcctcggacacttcttgaagcgaatgtgga 1906  
 Db 1921 GCCCTAACCAGGAGCCAGATTAAGAAACAGCTCGGCGCATTTTGAAGCCCAATGTGCA 1980  
 QY 1907 ggaagggagcagcagcagcgttggagagaagatcgaagatccagactctatccctt 1966  
 Db 1981 GGAAAGGAGAGCCAGCCGTTGGGAGAAATCTCAAGATCTCAAGATCTCAATCTCTTT 2040  
 QY 1967 cctctggccagtgaaattgtgtctcctcagccttggggagcctctctcttgaacctta 2026  
 Db 2041 CCTTGCCCAAGGAATTTGTGTCTTCCAGCTTTGGGAGCTCTTCTTGAACCTTGA 2100  
 QY 2027 taagaccacactgtagctctctctctcatccctctcctcctcctcctcctcctcctcctcct 2086  
 Db 2101 TAAGACCCACATGAGATCTCTCTCTCATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2160  
 QY 2087 tgcagagattgtcactcaaaccttactctgtgagctatcaataataaataacagatttatt 2146  
 Db 2161 TGCCAGAGATTTGATCTCAACCTTACTGTGAGCTCATTAATTAATTAACAGATTATT 2220  
 QY 2147 ttccagcttaaaaaa 2162  
 Db 2221 TTCCAGCTTAAAAAAA 2236

RESULT 3  
 LOCUS AB018687  
 DEFINITION Homo sapiens mRNA for autoimmune enteropathy-related antigen AIE-75, complete cds.  
 ACCESSION AB018687  
 VERSION AB018687.1 GI:5231270  
 KEYWORDS autoimmune enteropathy-related antigen AIE-75; AIE-75.  
 SOURCE Homo sapiens small intestine cDNA to mRNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (sites)  
 Kobayashi, I., Immura, K., Kubota, M., Ishikawa, S., Yamada, M., Tonoki, H., Okano, M., Storch, W. F., Moriuchi, T., Sakiyama, Y., and Kobayashi, K.  
 Identification of an autoimmune enteropathy-related 75-kilodalton antigen  
 JOURNAL Gastroenterology 117 (4), 823-830 (1999)  
 MEDLINE 99431904  
 REFERENCE 2 (bases 1 to 2213)  
 Kobayashi, I.  
 Direct Submission



|            |   |   |             |
|------------|---|---|-------------|
| QY         | 1614  | actgagatcaccctgtgtgtccgtctgcgccccaaggagatagcagatgagctgacct    | 1673        |
| Db         | 1681  | ACTGATGACACTGTGTGTGCGTGTGCCCCCAAGAGATATGACATGAGCTACCT         | 1740        |
| QY         | 1674  | tctgtctgaagtcacaaaggaggaacaaatctcaacgctgtaagaaacagtcagctcgccg | 1733        |
| Db         | 1741  | TCTT-CTGTAAGTCACAAAGGGGAAACCAATTCAC-CTTAGGAAACAGTGAAGCTCCGGC  | 1798        |
| QY         | 1734  | cccactctgtgaacacaaagcctcgagccagccttgagaggagagccatgacacacaca   | 1793        |
| Db         | 1799  | CCCACTGCTGAACACAAAGCCCTGGATACCTTGAAGAGGCCACATACACACACA        | 1858        |
| QY         | 1794  | gattgcatcctgtgagcctgaatctatcacccaggaatcacaactcctgtgcccctga    | 1853        |
| Db         | 1859  | GATGGCATCTTGGAGACCTGATCTATCACCCAGAGATCTCAACCTCCCTTGGCCCTGA    | 1918        |
| QY         | 1854  | accagggccagataaggaacagctcgggccactttttgaaggccaatgtgaggaagaag   | 1913        |
| Db         | 1919  | ACCAGGGCCAGATTAAGAACACCTCGGGCCACTCTTGAAGGCCAACGTGAGGAAGG      | 1978        |
| QY         | 1914  | gagcagccagccgcttttgaggaagatctcaagatccagactctatctctctctg       | 1973        |
| Db         | 1979  | GAGCAGCCAGCATTTGGAGAGATCTCAAGATCCAGACTCTCATCTTCTTCTG          | 2038        |
| QY         | 1974  | cccaatgaattgtctctcccaagctttggggagctcctcttgaaacctataagacc      | 2033        |
| Db         | 2039  | CCCAATGAATTTGGTCTCTCCAGCTCTGGGGAGCTCTTCTTGAACCTTAATAAGACC     | 2098        |
| QY         | 2034  | ccaatgaattgtctctctcccaagctttggggagctcctcttgaaacctataagacc     | 2093        |
| Db         | 2099  | CCACTGGAGTCTCTCTCTCCATCCCTCTCTGCCCCCTGCTTAATTCCTGCTGCAAG      | 2158        |
| QY         | 2094  | attgtcaactccaacttact 2114                                     |             |
| Db         | 2159  | ATTGTCACCTCCAAACCTTACT 2179                                   |             |
| RESULT 4   |   |   |             |
| AB006955   | 2283 bp   | mRNA  | 02-DEC-1999 |
| LOCUS      |   |   |             |
| DEFINITION | Homo sapiens mRNA for AIE-75, complete cds.                         |   |             |
| ACCESSION  | AB006955  |   |             |
| VERSION    | AB006955.1 GI:5152287   |   |             |
| KEYWORDS   | AIE-75; aie-75.   |   |             |
| SOURCE     | Homo sapiens small intestine cDNA to mRNA.                          |   |             |
| ORGANISM   | Homo sapiens  |   |             |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;       |   |             |
| AUTHORS    | Eutheria; Primates; Catarrhini; Hominiidae; Homo.                   |   |             |
|            | Kobayashi, I., Imamura, K., Kubota, M., Ishikawa, S., Yamada, M.,   |   |             |
|            | Tonoki, H., Okano, M., Storch, W.F., Moriuchi, T., Sakiyama, Y. and |   |             |
|            | Kobayashi, K.   |   |             |
| TITLE      | Identification of an autoimmune enteropathy-related 75-kilodalton   |   |             |
| JOURNAL    | antigen   |   |             |
| MEDLINE    | Gastroenterology 117 (4), 823-830 (1999)                            |   |             |
| REFERENCE  | 99431904  |   |             |
| AUTHORS    | 2 (bases 1 to 2283)   |   |             |
| TITLE      | Kobayashi, I.   |   |             |
| JOURNAL    | Direct Submission   |   |             |
|            | Submitted (31-AUG-1997) to the DDBJ/EMBL/GenBank databases. Ichiro  |   |             |
|            | Kobayashi, Hokkaido University School of Medicine, Department of    |   |             |
|            | Pediatrics, North-15, West-7, Kita-Ku, Sapporo, Hokkaido 060-8638,  |   |             |
|            | Japan (E-mail:ichikobae@med.hokudai.ac.jp, Tel:011-716-1161,        |   |             |
|            | Fax:011-736-9267)   |   |             |
| FEATURES   | Location/Qualifiers   |   |             |
| source     | 1..2283   |   |             |
|            | /organism="Homo sapiens"  |   |             |
|            | /db_xref="taxon:9606"   |   |             |
| gene       | /tissue_type="small intestine"                                      |   |             |
| CDS        | 246..1811   |   |             |
|            | /gene="aie-75"  |   |             |
|            | 246..1811   |   |             |

|    |     |  |     |
|----|-----|--|-----|
| QY | 131 | aggctgatttctgaatgaatgaatgaagaagagactatctctatgagtgtgcgaa        | 190 |
| Db | 187 | AGGTGATTTTCTGATTTGAATTAATGATGACAGAGAGACTATCTATGATGCTGCCGA      | 246 |
| QY | 191 | gttaccaacagacatgaatgagctgtgcgtctgtgtgagagactgaagctgtcatcaatg   | 250 |
| Db | 247 | TGTACACACAGACATGAGACGCGGCTGCTGTGGAGACCTGAAGCTGTGATCAATG        | 306 |
| QY | 251 | aaccagccgtctgtcctctgtttgtatgcacatcgccctgtatccactgaagaccaaag    | 310 |
| Db | 307 | AACCAAGCCGTGTGCTGTCTGTGATGCCATTCGGCCCTGATCCCATGAAGCACAGG       | 366 |
| QY | 311 | tggaaatgatcatgacgtgaccccccgcgctccaagaaagctgaaggaagtgcttgacc    | 370 |
| Db | 367 | TGGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG      | 426 |
| QY | 371 | gtctgacacccgaaagagctgtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt     | 430 |
| Db | 427 | GTTCTACCCCGAAGGCTCGGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT     | 486 |
| QY | 431 | tcttcattcccaactcaatcaagagcggtcaaggaagagcggtcggtccaggtagaag     | 490 |
| Db | 487 | TCTTATCTCCCACTCTCATCAAAAGCGGTCAAGACAGCCTCGGCTCCAGGTAGGGG       | 546 |
| QY | 491 | acgagatcgctccgagatcaatgatalctcatctcctctgtacccaatgaagagatcalca  | 550 |
| Db | 547 | ACGAGATCGTCGAGATCAATGATATTCATCTCTCTGATCCCATGAGAGGATCATCA       | 606 |
| QY | 551 | acctattcgcgaacaaagaactgtgtcattcaagaatgaagacatctggtcgtatccccg   | 610 |
| Db | 607 | ACCTATTTCGAACCAAGAAACCTGTGTCATCAAGTGAACATTCGCGCTGATCCCGG       | 666 |
| QY | 611 | tgaagaagctctccctgaatgaagccctcacttgagcagatgtagatgaagttgtgtgagat | 670 |
| Db | 667 | TGAAAAGCTCTCTGATATGAGCCCTCACTTGGCACTATGTGATCACTTTGTGTGGAAT     | 726 |
| QY | 671 | ctggggcgctgaggaagcagctgtggtctcccttgaaatcgggaaacaaggaagaagaag   | 730 |
| Db | 727 | CTGGGGCGTGGAGGAGCGCTGGGCTCCCTGGAATAAGGAAACAAGGAAGAAGAGG        | 786 |
| QY | 731 | tcttcatagcctgtgtaagctcccgagagcctgtgtgtaagatcttcacggcgcccatcc   | 790 |
| Db | 787 | TCTTCAATAGGCTGTGATGAGCTCCGAGGCTTGGCTGAGCATTTCCAGCGGCCCATACC    | 846 |
| QY | 791 | agaagccctgagatcttcatgagcattgaaacctgtgctcctctctcaggtgaggt       | 850 |
| Db | 847 | AGAAGCTTGGCATCTTTATCAGCATGTGAACCTGGCTCCCTGTCTGTAGGTGGAT        | 906 |
| QY | 851 | tggagatagggagacagatgtcgaaatcaatgagctgacttctcaacctgataca        | 910 |

Db 907 TGGAGATAGGGGACACGATTTGTGCAAGTCAATGCGCTGCACTTCTCACTGATGATCACA 966  
 QY 911 aggaagcttgtaaatgtgtctgtaaaaaatagccgagccctgaccattccattgtgtgctgag 970  
 Db 967 AGGAGGCTGTAAATGTCCTAAAGATGACCGGACGCTGACCCTTCATCTGATCTCTCAG 1026  
 QY 971 ctggccgggagctgttcatgtacagaccgggagcgctgtgcgaagcgcgagcgctgtgagc 1030  
 Db 1027 CTGGCCGGGAGCTTTCATGTACACAGCCGGGAGCGGCTGACAGAGGCCGGCAGGCTAGC 1086  
 QY 1031 tgcagcggcagagagcttcatgtcgaagaagcgctgtgcagtagagttccaaagaatcc 1090  
 Db 1087 TGCAGCGGAGAGAGCTTCTCATGCAAGAGCGGCTGGCGATGAGAGTCCAAACAAGTCTCTCC 1146  
 QY 1091 aggaagcagcagagatgtgagcgcgcaaaagagaaataatgtcccaagaagcagcagag 1150  
 Db 1147 AGGAGCGAGAGAGATGAGATGAGCGGCAAGGAGAAATTTGCCCAAGGACGACAGAGG 1206  
 QY 1151 aaaaatgagataaccggaagagatgtgacaagattgtagaaggaagagaagattaga 1210  
 Db 1207 AAATATGAGATATACCGGAGAGAGATGACAGATTTGAGAGGAGAGAGAGATTTTAA 1266  
 QY 1211 agcaatggaagaagagactggggtctcaaaagaaacagctactcttgcctaaacacactc 1270  
 Db 1267 AGCAATGGAAGAAAGACTGGGGCTCAAAAGAACAGTACTCTTGTCTTAAACCATCACTG 1326  
 QY 1271 ctgaagtgacacccagtaaccctctgcgaagcgaagt----- 1307  
 Db 1327 CTGAGGTACACCGGTACCCCTTCGCAAGCGCAAGATATGATCAGGAGTGAGAACCTGAGC 1386  
 QY 1307 -----gattcggga 1316  
 Db 1387 TCGAGCCCGAGATGACCTGATGAGAGGACGAGGAGAGAGACAGATTTCCGGA 1446  
 QY 1317 aatatgagaagagcttgcacccctactactatgttaccaccaagcagatcatcagggagaag 1376  
 Db 1447 AATATGAGGAAGGCTTTGAGCCCTACTACTATGTTGTTACCCAGAGACAGATCATGGGGAGG 1506  
 QY 1377 atgtcgggtctctacgcatcaagaagaagagatccttagaacctgcccctggaagcggtg 1436  
 Db 1507 ATGTCGGGCTTCATGCGATCAAGAGAGGAGATCCTTAAACCTGGCCCTGGAAGCGCGTG 1566  
 QY 1437 tggactcccccattgggaagagtggtgttctgtctgtctatgagcgggagctgtgagc 1496  
 Db 1567 TGGACTCCCCATTGGGAAGGTGCTCTTTCTGTGTATGACGGGGAGCTGCTAGC 1626  
 QY 1497 ggcattgtgcattgtgaaagggaagcagatcatcgtcaatcaacgcaagatgtgacag 1556  
 Db 1627 GGCATGCTGCAATTGTGAAGGGGAGAGATCATGCGCAATCAACGGCAAGTTGTGACAG 1686  
 QY 1557 actaacctctgctgagctgagctgagcgtgcctgcgaagagcctggaaatcagggcgagact 1616  
 Db 1687 ACTACACCCCTGCGTGAAGGCTGAGGCTGCCCTGCAGAGAGCGCTGGAATCAGGGCGGACT 1746  
 QY 1617 ggcattgacactgt 1676  
 Db 1747 GGAATCGACTTGT 1806  
 QY 1677 tgcgtgaagtcacaaagggaacccaatltcacgctgttagaagaagctgagctccgagccc 1736  
 Db 1807 T-CTGAAGTCCAAAGGGGAAACCAATTCAC- CGTTAGGAACAGTGAAGCTCCGGGCC 1864  
 QY 1737 acctgtgaaacacaaagcctgcgagcagccttgaagaagggccacatgacacacacagat 1796  
 Db 1865 ACCTCGTGAACACAAACCCCTGAGTACAGCCTTGAAGAGGCGCACACTACACACACAGAT 1924  
 QY 1797 ggcattcttgagccttaactatcaccaggaatccaactcccttggccttgagcc 1856  
 Db 1925 GGCATCTCTTGGGACTTAATCTATCACCAAGAAATCTCAAACTCTTTGGCCCTTGAC 1964  
 QY 1857 agggccagataaagaagcagctcgagccactttttgaagggccaatgtcgagaagaaaggag 1916  
 Db 1985 AGGGCCAGATTAAGGAAGCTCGGGCCACTCTTGTGAAGGCCAAACGTGGAAGAAAGGAG 2044

QY 1917 cagccagccgcttggggaagaatctcaagatccagatccagatccatctcttcccttgagcc 1976  
 Db 2045 CAGCCAGCCCAATTTGGGAGAAATCTCAAGATCCAGACTCTCATCTCTTCTGTGGCC 2104  
 QY 1977 agtgaattgtctctccacgcttgggagctcctctccttgaacccaataagaccaca 2036  
 Db 2105 AGTGAATTTGGTCTCTCCAGCTGTGGGGA-TCCTTCTTTGAACCTTAATGAACCCCA 2163  
 QY 2037 ctgagatctctctctccatccctctctctctgtccctctgtcctaattgtgcagagatt 2096  
 Db 2164 CTGAGTCTCTCTCTCAATCCCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2223  
 QY 2097 gtcaatccaaacttactctgagctcatataaataaagaattatttccagctta 2156  
 Db 2224 GTCACTCCAAACCTTACT-----GCAGCTCATCGAAATAATTAATTAACCACTAA 2277  
 QY 2157 aaaaaa 2162  
 Db 2278 AAAAAA 2283

RESULT 5  
 AC005137/c  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

Submitted (30-SEP-1998) Genome Science & Technology Center,  
 University of Texas Southwestern Medical Center, 5323 Harry Hines  
 Blvd, Dallas, TX 75235-8591, USA  
 On Oct 1, 1998 this sequence version replaced gi:3241950.  
 Further information regarding the map of this region or  
 annotation of 6-106f23 can be found at  
 http://gestec.swmed.edu/chromosome2.htm.  
 CHROMOSOMAL LOCUS: This PAC clone comes from the Usher syndrome  
 region (USH1C) mapped between STS markers D11S1310 and D11S1414.  
 MARKER CONFIRMATION: STS/ESTs sequence confirmed; D11S1032 and  
 D11S1438.  
 MAPPED CLONE OVERLAP: pD2339b2 and 6-109h6.



IMPORTANT: This submission contains the entire insert of clone 6-106f23. 6-106f23 comes from a PAC library constructed at the Roswell Park Cancer Institute by the Pictet de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.40. In addition, this sequence has been finished such that 99.9% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

## FEATURES

```
Source      Location/Qualifiers
repeat_region 1..161190
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              574..955
              /rpt_family="MT2B2"
repeat_region 1445..1635
              /rpt_family="MIR"
              3125..3302
              /rpt_family="MIR"
              11175..11329
              /rpt_family="MIR"
              14552..14716
              /rpt_family="MIR"
              complement(18589..
              23165..23232)
              /rpt_family="MIR"
              complement(25103..
              29622..29909)
              /rpt_family="Alu"
              complement(32772..
              33175..33284)
              /rpt_family="MIR"
              complement(34762..
              34864..34996)
              /rpt_family="MIR"
              complement(35719..
              36817..37084)
              /rpt_family="MIR"
              complement(37953..
              39000..39177)
              /rpt_family="MIR"
              complement(39898..
              41579..41780)
              /rpt_family="MIR"
              complement(43317..
              45609..45899)
              /rpt_family="Alu"
              complement(46819..
              48440..48895)
              /rpt_family="MIR"
              complement(50385..
              50842..51126)
              /rpt_family="Alu"
              complement(53140..
              55758..56060)
              /rpt_family="Alu"
              complement(59724..
              66550..66645)
```

```
repeat_region 73180..73260
              /rpt_family="Alu"
              complement(74694..
              77037)
              /rpt_family="MIR"
              complement(77130..
              77414)
              /rpt_family="MIR"
              complement(77532..
              77641)
              /rpt_family="MIR"
              complement(78518..
              78630)
              /rpt_family="Alu"
              complement(78918..
              85547..85650)
              /rpt_family="MIR"
              complement(88613..
              88499..88613)
              /rpt_family="MIR"
              complement(93509..
              93900)
              /rpt_family="Alu"
              complement(94148..
              95822)
              /rpt_family="Alu"
              complement(95536..
              96013..96567)
              /rpt_family="Alu"
              complement(106391..
              114951..115136)
              /rpt_family="MIR"
              complement(118500..
              118580)
              /rpt_family="MIR"
              complement(126403..
              126593)
              /rpt_family="Alu"
              complement(126942..
              127352)
              /rpt_family="Alu"
              complement(128074..
              128639)
              /rpt_family="Alu"
              complement(128923..
              130009..130278)
              /rpt_family="Alu"
              complement(134144..
              137205..137469)
              /rpt_family="Alu"
              complement(140667..
              142521)
              /rpt_family="MIR"
              complement(148006..
              149821..150066)
              /rpt_family="MIR"
              complement(151137..
              160309..160771)
              /rpt_family="MIR"
              complement(161190..
              161190)
```

## BASE COUNT

37674 a 40113 c 40195 g 43208 t

## ORIGIN

Query Match 20.2% Score 437.4; DB 11; Length 161190;  
Best Local Similarity 95.7%; Pred.No. 9.9e-91;  
Matches 472; Conservativity 0; Mismatches 16; Indels 5; Gaps 2;

QY 1666 gctgaacctctgtcgaagtcgaaggaaggaacaaatcgcggttaggaagaacagtga 1725  
DB 5253 GCAGACCTTCTTGCAAGTCCAAAAGGGAACCAATTCAC-CGTTAGGAACAGTGA 5195



QY 1726 gctcggccccacatcgtgacacaaagcctcgaccacatcgtgaggaagccacatgac 1785  
|||||  
Db 5194 gctccggccccacatcgtgacacaaagcctcgaccacatcgtgaggaagccacatgac 5135  
|||||  
QY 1786 acacacagatgacatccttggagctgaaatctatcacccaggaatctcaactccctt 1845  
|||||  
Db 5134 acacacagatgacatccttggagctgaaatctatcacccaggaatctcaactccctt 5075  
|||||  
QY 1846 ggcctcgacacagggccaggaatgaagaaacagctgggacacatcttttgaagccaatgtg 1905  
|||||  
Db 5074 ggcctcgacacagggccaggaatgaagaaacagctgggacacatcttctgaagccaatgtg 5015  
|||||  
QY 1906 aggaagagggagccagccagctgttggaagaagatctcaagatccagacatccttccct 1965  
|||||  
Db 5014 aggaagagggagccagccagctgttggaagaagatctcaagatccagacatccttccct 1955  
|||||  
QY 1966 tccctgtgcccagtgaaattgtgtctctccagctgttggaagacatccttccctgaacccta 2025  
|||||  
Db 4954 tccctgtgcccagtgaaattgtgtctctccagctgttggaagacatccttccctgaacccta 4895  
|||||  
QY 2026 ataaagacccacatgagctctctctctccatccctctctctgcccctgctcctaattg 2085  
|||||  
Db 4894 ataaagacccacatgagctctctctctccatccctctctctgcccctgctcctaattg 4835  
|||||  
QY 2086 ctgcagagatgtcactccaaccttactctgagctcatatataaataacagattat 2145  
|||||  
Db 4834 ctgcagagatgtcactccaaccttactctgagctcatatataaataacagattat 4779  
|||||  
QY 2146 ttccacagctaa 2158  
|||||  
Db 4778 ttccacagctaa 4766  
|||||  
RESULT 6  
LOCUS G55504 463 bp DNA STS 26-OCT-1999  
DEFINITION SHGC-100852 Human Homo sapiens STS genomic, sequence tagged site.  
ACCESSION G55504  
VERSION G55504.1 GI:6120823  
KEYWORDS STS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 463)  
AUTHORS Myers,R.M.  
TITLE Human STS (1999)  
JOURNAL Unpublished (1999)  
COMMENT  
Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, M-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu  
Primer A: TCAGAGTCAGGAATTCGACATCT  
Primer B: TGTAAGGGAACGAGATCATCG  
STS size: 313  
PCR Profile:  
Initial incubation: 95 degrees C for 10 minutes  
Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700  
Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
AmpliTaq Gold Polymerase: 0.07 units/ul  
Total Vol: 5 ul

Buffer: MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3  
BAC ends sequenced at TIGR from the RPO11 BAC library. Designed and developed at the Stanford Human Genome Center.  
FEATURES  
source  
1..463  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="11"  
/clone\_lib="Human"  
STS  
primer\_bind 80..392  
primer\_bind 80..102  
BASE COUNT complement(370..392)  
96 a 116 c 128 g 123 t  
ORIGIN  
Query Match 4.9%; Score 105; DB 13; Length 463;  
Best Local Similarity 82.8%; Pred. No. 4.5e-14;  
Matches 120; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 1502 ggtggcatttgaaagggagcagatcatatgcaatcaacggcaagattgtgacagactac 1561  
|||||  
Db 400 ggtggcatttgaaagggagcagatcatatgcaatcaacggcaagattgtgacagactac 341  
|||||  
QY 1562 accctgtgagctgagctgagctgagctgagctgagctgagctgagctgagctgagctgagct 1621  
|||||  
Db 340 accctgtgagctgagctgagctgagctgagctgagctgagctgagctgagctgagctgagct 281  
|||||  
QY 1622 gacctgtgtgtccgctgtccccc 1646  
- - - - -  
Db 280 AGGCCCTCCCTCCTTCCTCCCTCC 256  
- - - - -  
RESULT 7  
LOCUS 166494 7218 bp DNA PAT 28-DEC-1997  
DEFINITION Sequence 14 from patent US 5670367.  
ACCESSION 166494  
VERSION 166494.1 GI:2724471  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 7218)  
AUTHORS Dornier,F., Scheifflinger,F. and Palmer,F. Gunter.  
TITLE Recombinant fowlpox virus  
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;  
FEATURES  
source  
1..7218  
/organism="unknown"  
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others  
ORIGIN  
Query Match 3.5%; Score 75.8; DB 5; Length 7218;  
Best Local Similarity 6.1%; Pred. No. 3e-07; Mismatches 157; Indels 0; Gaps 0;  
Matches 26; Conservative 240; Mismatches 157; Indels 0; Gaps 0;  
QY 829 tccctgtcgtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 888  
|||||  
Db 1478 TACCTATCTATGCAAGAGCTTAAGAGATGAAAGAAATTGTGCTACRRRRRRRRRRRR 1419  
|||||  
QY 889 gacctctaacctgagtcacaaaggagctgtaaatgtctgaaaaaataagccagcgtg 948  
- - - - -  
Db 1418 RRR 1359  
- - - - -  
QY 949 accatcccaattgtagtgagctgagcggagagctgtctacagacacggagcgagctg 1008  
- - - - -

[illegible]

| TITLE         | COMMENT   | FEATURES   |
|---------------|---|--|
| JOURNAL       | Direct Submission<br>Submitted (30-JUN-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA<br>On Jun 4, 1999 this sequence version replaced gi:4966400.<br>All repeats were identified using RepeatMasker: Smlt, A.F.A. & Green, P. (1996-1997)<br><a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> | Only 180385 base pairs from the middle of this clone are being submitted. The remainder overlaps either accession number AC00515 (WICGR project L335) or accession number AC005243 (WICGR project L343). |
| source        | Location/Qualifiers   |  |
| repeat_region | 1. 180385   |  |
| repeat_region | /organism="Homo sapiens"  |  |
| repeat_region | /db_xref="taxon:9606"   |  |
| repeat_region | /clone="84.E.24"  |  |
| repeat_region | /clone_1lb="Alan Buckler -- per comm"   |  |
| repeat_region | /map="17"   |  |
| repeat_region | /chromosome="17"  |  |
| repeat_region | 446. .753   |  |
| repeat_region | /rpt_family="Alusc"   |  |
| repeat_region | 2375. .2419   |  |
| repeat_region | /rpt_family="TTTGG)n"   |  |
| repeat_region | complement(2985. .3103)   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | complement(3146. .3352)   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | complement(3531. .3639)   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | complement(3798. .4394)   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | 5531. .5631   |  |
| repeat_region | /rpt_family="MER81"   |  |
| repeat_region | 5648. .5674   |  |
| repeat_region | /rpt_family="(CAAAA)n"  |  |
| repeat_region | 5812. .6020   |  |
| repeat_region | /rpt_family="L1MC4"   |  |
| repeat_region | 6180. .6224   |  |
| repeat_region | /rpt_family="AT_rich"   |  |
| repeat_region | 6303. .6611   |  |
| repeat_region | /rpt_family="Alusp"   |  |
| repeat_region | 7378. .7618   |  |
| repeat_region | /rpt_family="MIR"   |  |
| repeat_region | 9593. .9623   |  |
| repeat_region | /rpt_family="(GGGA)n"   |  |
| repeat_region | complement(10666. .10923)   |  |
| repeat_region | /rpt_family="Alusq"   |  |
| repeat_region | 10958. .11088   |  |
| repeat_region | /rpt_family="MIR"   |  |
| repeat_region | 11752. .11914   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | complement(11915. .12219)   |  |
| repeat_region | /rpt_family="Alusx"   |  |
| repeat_region | 12220. .12347   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | 13110. .13182   |  |
| repeat_region | /rpt_family="GA-rich"   |  |
| repeat_region | 13526. .13564   |  |
| repeat_region | /rpt_family="(TCC)n"  |  |
| repeat_region | 13575. .13660   |  |
| repeat_region | /rpt_family="(TA)n"   |  |
| repeat_region | complement(14380. .14658)   |  |
| repeat_region | /rpt_family="AluJo"   |  |
| repeat_region | 14742. .14830   |  |
| repeat_region | /rpt_family="(TA)n"   |  |
| repeat_region | 14860. .14967   |  |
| repeat_region | /rpt_family="(TA)n"   |  |
| repeat_region | complement(14977. .15122)   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | complement(15292. .15385)   |  |
| repeat_region | /rpt_family="L2"  |  |
| repeat_region | 16672. .16781   |  |







SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Homidae; Homo.  
TITLE 1 (bases 1 to 195385)  
JOURNAL Genome Therapeutics Corporation Sequencing Center: Human Genome  
REFERENCE Unpublished  
AUTHORS Smith,D.R.  
TITLE 2 (bases 1 to 195385)  
JOURNAL Direct Submission  
REFERENCE Submitted (08-DEC-1999) Genome Therapeutics Corporation, 100 Beaver  
AUTHORS Street, Maitland, MA 02453, USA  
COMMENT \* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 48 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1 1112: contig of 1112 bp in length  
\* 1113: gap of unknown length  
\* 2221: contig of 1109 bp in length  
\* 2222: gap of unknown length  
\* 3230: contig of 1009 bp in length  
\* 4630: gap of unknown length  
\* 3231: contig of 1400 bp in length  
\* 4631: gap of unknown length  
\* 5801: contig of 1170 bp in length  
\* 5801: gap of unknown length  
\* 7675: contig of 1875 bp in length  
\* 7676: gap of unknown length  
\* 9606: contig of 1931 bp in length  
\* 9607: gap of unknown length  
\* 10867: contig of 1261 bp in length  
\* 12422: gap of unknown length  
\* 10868: contig of 1555 bp in length  
\* 12423: gap of unknown length  
\* 13550: contig of 1128 bp in length  
\* 13551: gap of unknown length  
\* 14811: contig of 1261 bp in length  
\* 15980: gap of unknown length  
\* 14812: contig of 1169 bp in length  
\* 15981: gap of unknown length  
\* 18262: contig of 2282 bp in length  
\* 18263: gap of unknown length  
\* 19306: contig of 1044 bp in length  
\* 19307: gap of unknown length  
\* 21169: contig of 1863 bp in length  
\* 21170: gap of unknown length  
\* 22313: contig of 1144 bp in length  
\* 22314: gap of unknown length  
\* 23697: contig of 1384 bp in length  
\* 23698: gap of unknown length  
\* 25371: contig of 1674 bp in length  
\* 25372: gap of unknown length  
\* 26897: contig of 1526 bp in length  
\* 26898: gap of unknown length  
\* 28765: contig of 1868 bp in length  
\* 28766: gap of unknown length  
\* 32095: contig of 3330 bp in length  
\* 32096: gap of unknown length  
\* 34175: contig of 2080 bp in length  
\* 34176: gap of unknown length  
\* 36174: contig of 1999 bp in length  
\* 36175: gap of unknown length  
\* 38196: contig of 2022 bp in length  
\* 38197: gap of unknown length  
\* 40333: contig of 2137 bp in length  
\* 40334: gap of unknown length  
\* 42398: contig of 2065 bp in length

\* 42399: gap of unknown length  
\* 44266: contig of 1868 bp in length  
\* 44267: gap of unknown length  
\* 47804: contig of 3558 bp in length  
\* 47805: gap of unknown length  
\* 51088: contig of 3284 bp in length  
\* 51089: gap of unknown length  
\* 55348: contig of 4260 bp in length  
\* 55349: gap of unknown length  
\* 58075: contig of 2727 bp in length  
\* 58076: gap of unknown length  
\* 61009: contig of 2934 bp in length  
\* 61010: gap of unknown length  
\* 64859: contig of 3850 bp in length  
\* 64860: gap of unknown length  
\* 68394: contig of 3535 bp in length  
\* 68395: gap of unknown length  
\* 74770: contig of 6376 bp in length  
\* 74771: gap of unknown length  
\* 78978: contig of 4208 bp in length  
\* 78979: gap of unknown length  
\* 85405: contig of 6427 bp in length  
\* 85406: gap of unknown length  
\* 90787: contig of 5382 bp in length  
\* 90788: gap of unknown length  
\* 97005: contig of 6218 bp in length  
\* 97006: gap of unknown length  
\* 102331: contig of 5326 bp in length  
\* 102332: gap of unknown length  
\* 109949: contig of 7618 bp in length  
\* 109950: gap of unknown length  
\* 118555: contig of 8606 bp in length  
\* 118556: gap of unknown length  
\* 126644: contig of 8089 bp in length  
\* 126645: gap of unknown length  
\* 136384: contig of 9740 bp in length  
\* 136385: gap of unknown length  
\* 149159: contig of 12775 bp in length  
\* 149160: gap of unknown length  
\* 162077: contig of 12918 bp in length  
\* 162078: gap of unknown length  
\* 181175: contig of 19098 bp in length  
\* 181176: gap of unknown length  
\* 195385: contig of 14210 bp in length.  
1. 195385  
Location/Qualifiers  
FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-175M21"  
BASE COUNT 56722 a 41286 c 40797 g 56532 t 48 others  
ORIGIN  
Query Match 2.9%; Score 63.4; DB 43; Length 195385;  
Best Local Similarity 75.2%; Pred. No. 0.00028;  
Matches 79; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 854 agataggagacgaattgtcgaatgacatgagctgactctactacatgacacag 913  
|||||  
DB 12775 AGATAGGGGACCGAATGTGCGAATGCGGTGACCTCTTAACCTGATCAAGG 12834  
QY 914 aggcgttaattgtcgaataaataagcgcgaacatcacatccca 958  
||| | |||  
DB 12835 AGGTAGATGTGGGGCTTCAACCTGTGGCCCTTGTATCTCA 12879  
RESULT 14  
AF148805/c DNA VRL 02-AUG-1999  
LOCUS  
DEFINITION  
AF148805 28559 bp  
Kaposi's sarcoma-associated herpesvirus ORF 68 gene, partial cds;  
and ORF 69, Kaposin, v-FLIP, v-cyclin, latent nuclear antigen, ORF  
K14, v-GPCR, putative phospholipase for myo-inositol synthase,  
and LAMP (LAMP) genes, complete cds.

|               |   |
|---------------|---|
| ACCESSION     | AF148805  |
| VERSION       | AF148805.1  |
| KEYWORDS      | GI:5669888  |
| SOURCE        | Kaposi's sarcoma-associated herpesvirus - Human herpesvirus 8.  |
| ORGANISM      | Kaposi's sarcoma-associated herpesvirus<br>viruses; dsDNA viruses, no RNA stage; Herpesviridae;<br>Gammaherpesvirinae; Rhadinovirus.  |
| REFERENCE     | 1 (bases 1 to 28559)  |
| AUTHORS       | Glenn, M., Rainbow, L., Aurad, F., Davison, A. and Schulz, T.F.   |
| TITLE         | Identification of a spliced gene from Kaposi's sarcoma-associated<br>herpesvirus encoding a protein with similarities to latent membrane<br>proteins 1 and 2a of Epstein-Barr virus   |
| JOURNAL       | J. Virol. 73 (8), 6953-6963 (1999)  |
| MEDLINE       | 99329221  |
| REFERENCE     | 2 (bases 1 to 28559)  |
| AUTHORS       | Glenn, M.A., Rainbow, L., Aurade, F., Davison, A.J. and Schulz, T.F.  |
| TITLE         | Direct Submission   |
| JOURNAL       | Submitted (15-Apr-1999) MRC Virology Unit, Church Street, Glasgow<br>G11 5UR, UK  |
| FEATURES      | Location/Qualifiers   |
| source        | 1..28559  |
|               | /organism="Kaposi's sarcoma-associated herpesvirus"   |
|               | /isolate="GK18"   |
|               | /db_xref="taxon:37296"  |
|               | /note="right end of the unique region plus 8 complete<br>copies and 1 partial copy of the terminal repeat (TR)"   |
| gene          | <1..2105  |
|               | /gene="ORF 68"  |
|               | <1..1102  |
| CDS           | /gene="ORF 68"  |
|               | /function="involved in DNA packaging"   |
|               | /note="similar to HSV-1 UL32, HCMV UL52 and EBV BFLF1"  |
|               | /codon_start=2  |
|               | /evidence=not_experimental  |
|               | /protein_id="A046496.1"   |
|               | /db_xref="GI:5669889"   |
|               | /translation="IMHLPPEEATYGLFGGRLMGIDQLHPEVOKCFKTAEKI<br>LGISNLOPKSEFIRGMLTGITTCNCFKTSNPRIDKEATGPTCCOITTTTPAS<br>GIPELARFCGASRPKPSLPLALIDWTSSELDPRPRPLASDSEKSVASH<br>DPEFSPLOADTSQPCLMHPPLGLRYNGTASVCLCECLAAHEAKALQTLCEV<br>MGIENNKKLVRIAEVLDNPFAMPYVSDPLRELIRCTQEIHKHLFCPLCALNA<br>KVSIEDVFLRPREDEYKKLRASAAAGOLDANTLFCEVVQTLVFLFKGQNARVGK<br>TTSLDIRELTAAQLKRRHLRLAHSQTSHTYA" |
| gene          | 1135..2105  |
|               | /gene="ORF 69"  |
| CDS           | 1135..2043  |
|               | /gene="ORF 69"  |
|               | /note="similar to HSV-1 UL31, HCMV UL53 and EBV BFLF2"  |
|               | /codon_start=1  |
|               | /evidence=not_experimental  |
|               | /protein_id="A046497.1"   |
|               | /db_xref="GI:5669890"   |
|               | /translation="MPKSVSSHISLATSTGRSGPRDIRCLSSRLRVPGARSAVS<br>SKHRGKRFISDKVFEFSILSHRELIGVDFLEMETPTCTVWLPLDLSTVAGRC<br>VLSRFGSSNMGQCACPTENETVAGSGSPQTMGDALKNNELCSVALAFIHHH<br>DKVIOHKTFYLSLSHSDVVRQSTLQGLTANVLKTFGHDPLPTITNGMLTMC<br>ILFTRALHGETALRLMDNLPNKISADCCROSSYVKEFVPTHDTSTIAVQHTIC<br>EVAVALDCTDEMRDIOKGTALVNAL"   |
| polyA_signal  | 2100..2105  |
|               | /gene="ORF 69"  |
|               | /note="possible signal for ORFs 68 and 69"  |
|               | /evidence=not_experimental  |
| polyA_site    | complement(2126)  |
|               | /gene="ORF K12"   |
|               | /evidence=experimental  |
| mRNA          | complement(2126..2824)  |
|               | /gene="ORF K12"   |
|               | /product="T0.7"   |
|               | /evidence=experimental  |
| gene          | complement(2126..2824)  |
|               | /gene="ORF K12"   |
|               | complement(2139..2144)  |
| polyA_signal  | /gene="ORF K12"   |
|               | /evidence=experimental  |
|               | complement(2616..2798)  |
| CDS           | /gene="ORF K12"   |
|               | /note="transforming gene"   |
|               | /codon_start=1  |
|               | /evidence=experimental  |
|               | /product="Kaposin"  |
|               | /protein_id="A046499.1"   |
|               | /db_xref="GI:5669892"   |
|               | /translation="MDKGLTVFAVHPVDLLNGMRWRLGAIPLVLCILASVPPSG<br>ORGPVAFRTYATGANH"   |
| 5'UTR         | complement(2799..2824)  |
|               | /gene="ORF K12"   |
|               | /evidence=experimental  |
| repeat_region | 2820..3081  |
|               | /rpt_type=direct  |
|               | /rpt_unit=2820..2842  |
|               | 3082..3436  |
| repeat_region | /rpt_type=direct  |
|               | /rpt_unit=3084..3106  |
|               | complement(6928..6933)  |
| polyA_signal  | /gene="ORF 71"  |
|               | /note="possible signal for ORF 71"  |
|               | /evidence=not_experimental  |
| gene          | complement(6928..7550)  |
|               | /gene="ORF 71"  |
|               | complement(6984..7550)  |
| CDS           | /gene="ORF 71"  |
|               | /function="inhibitor of apoptosis"  |
|               | /note="contains two death effector domains"   |
|               | /codon_start=1  |
|               | /evidence=experimental  |
|               | /product="v-Flip"   |
|               | /protein_id="A046498.1"   |
|               | /db_xref="GI:5669891"   |
|               | /translation="MATVEVCEYARKLGTDDREYVFLNWFIPQPTLAOLIGLARA<br>LKEGRLEPLLAECLEFRAGRRDLRLDLHDPRLTEGLAGTSYSPYQTYIAYD<br>GELCARLRSILFLSKDTIGSRSTQPTLHWYTCMENLDLGPDDVDALMSLNSLR<br>VDLQVQVTLKGLHLSGSHQHRHTP"  |
| polyA_signal  | complement(7601..7606)  |
|               | /gene="ORF 72"  |
|               | /note="possible signal for ORFs 72 and 73"  |
|               | /evidence=not_experimental  |
| gene          | complement(7601..8406)  |
|               | /gene="ORF 72"  |
|               | complement(7601..12037)   |
| gene          | /gene="ORF 73"  |
|               | complement(7633..8406)  |
| CDS           | /gene="ORF 72"  |
|               | /note="cyclin D homologue"  |
|               | /codon_start=1  |
|               | /evidence=experimental  |
|               | /product="v-cyclin"   |
|               | /protein_id="A046500.1"   |
|               | /db_xref="GI:5669893"   |
|               | /translation="WATANNPSSGLDPTCEDRIYVILIEPRELTSDSVGTQQ<br>SLTSHMRLLGTWMSVCOEYNEBNVVALMLDLRLIKVSKERFQTSACIL<br>VASKRLSTPSTSGYAAADSFSGRLIDQEKRLIARFPAVATVTSPLIL<br>KLGGSHLDPMHHEVNTLITKALVDPRTGSLPASIIAAGCALLVPANVLPDTHSG<br>GVVQLASITICDVSVAQAVQILTSVSDRLRLDSY"  |
| CDS           | complement(8648..12037)   |
|               | /gene="ORF 73"  |
|               | /note="LNA: component of latency-associated nuclear<br>antigen: LANA"   |
|               | /codon_start=1  |
|               | /evidence=experimental  |
|               | /product="latent nuclear antigen"   |
|               | /protein_id="A046501.1"   |
|               | /db_xref="GI:5669894"   |
|               | /translation="WAPPGMRLSRSGRTGAPLTRGSCRRNRSPERCDLGDLHQPR<br>RKHVADSVGREGCPHTLPITGSPVTFSGLPATVSPPLPAPATPLPPPA<br>LLPVTSSSPITPSHPVSGTQTHNSPALPTQSPRESSORPLSSPGRDPSRP<br>MRPPSQQTTPPHSPPTTPPPPPSKSSPDSIAPSTIRSLRRRLSSPQSPSTLNPICO   |







**THIS PAGE BLANK (USPTO)**